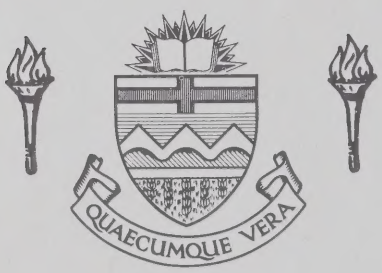

THE UNIVERSITY OF ALBERTA

ESTIMATION OF QUANTILE IN FINITE POPULATION WITH SUPER-POPULATION MODEL

by

MD. ZAHORUL ISLAM

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES AND RESEARCH

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE

OF MASTER OF SCIENCE

DEPARTMENT OF

STATISTICS AND APPLIED PROBABILITY

EDMONTON, ALBERTA

SPRING, 1983

ABSTRACT

A literature survey on various types of super-population models and their uses in finite population sampling is given. Topics like optimum sampling, balanced samples and randomization in survey sampling are discussed in detail. We have proved some properties of univariate distribution of quantile of a sample from a finite population. One of our main objectives was to explore the possible use of auxiliary informations for estimating finite population quantiles. Keeping this goal in our mind we have derived the bivariate distribution of smaple quantiles and their asymptotic distribution. Assuming a certain super-population model we have suggested an estimator for finite population quantiles which involves the auxiliary variable. Further study will be required to establish useful properties of the suggested estimator.

ACKNOWLEDGEMENT

I wish to express my grateful thanks to my supervisor, Professor K.L. Mehra for introducing me to the problem and for the considerable time and effort that he has spent on my work in preparing this thesis.

I would also like to thank Mr. Andrew Luong and Ms. Martha S. Rhodes for helping me in proofreading and M. June Talpash for her excellent typing of this thesis.

TABLE OF CONTENTS

CHAPTER I

INTRODUCTION

In sample survey theory the basic assumption is that we have a fixed finite population of N identifiable units under study. Due to lack of time and resources which includes money, expertise, etc., we are constrained to study only a part of the population concerned. Of course, there are situations where a complete study of the population is not at all feasible and one has to depend on sample survey methods. The objective in survey sampling is to make inference about some characteristics of the population. Most of the literature on sample surveys deals with the estimation of the population total, population mean, population proportion and standard errors of their estimates.

The object of this thesis is to study recent developments in survey sampling dealing with the estimation of quantiles of finite population and inference under super population models. It is a common practice to use auxiliary information for estimating population mean or total. Analogously, our interest in this work is to study the use of auxiliary information for improving the estimation of finite population quantiles. For example, can we use with benefit the information on quantiles of auxiliary variables in estimating the quantile of the main variables? Keeping this objective in mind we have derived the bivariate distribution of sample quantiles and studied its asymptotic behaviour.

In this chapter we shall try to point out some shortcomings of the conventinal approach and a brief history on the development of

the model based approach to inference in survey sampling as an alternative to the conventional one.

Let there be $N < \infty$ units in the population. N is called the population size. The units are identifiable, that is units of the population can be uniquely labelled from 1 to N and the label of each unit is known. We can denote the population by $\mathcal{U} = \{1,2,\ldots,N\}$. With each unit i, there is associated a measurement y_i on a variable character y. For all practical purposes y_i is real for all $i \in \mathcal{U}$. We shall represent the auxiliary information on U_i by a real measurement x_i or by vector $\underset{\sim}{x}_i$. Our target is to estimate the population total, $y = \sum_{i}^{N} y_i$. The method is to draw a representative sample of the population and on the basis of sample observations we have to estimate population total. Let $s = \{i_1,\ldots,i_n\}$ be sampled units without repetition and $\bar{s} = \mathcal{U} - s = \mathcal{U} - \{i_1,\ldots,i_n\}$ be units not in the sample s. Then, we have population total,

$$y = \sum_{i \in s} y_i + \sum_{i \in \bar{s}} y_i \quad . \tag{1.1}$$

In estimating y, the first sum on the right hand side of (1.1) is exactly known to us (assuming there is no measurement error) and our object is to estimate the second sum, namely the total of the non-sampled part of the population. Survey methodology available in common survey sampling textbooks, (e.g. Cochran, 1977) is devoted primarily to finding a good survey design, suitable to the practical situtation for estimating the unknown part of the population total. This approach is now commonly known as the conventional approach or the design based approach. At this point it is interesting to mention a few lines from Basu (1969):

"The objective of planning a survey should be to end up with a good sample. The term 'representative sample' has been useed in survey terminology. But no one has cared to give a precise definition of the term. It is implicitly taken for granted that statistician with his biased mind is unable to select a representative sample. So a simplistic solution is sought by turning to an unbiased die (the random number tables). Thus, a deaf and dumb die is supposed to do the job of selecting a 'representative sample' better than a trained statistician."

Broadly speaking, there are three main methods of estimation of finite population parameters. These are: Methods based on

1. measurements of units which are exact, that is, there is no error in measurements;
2. measurements which are not exact but subject to random errors; and
3. knowledge of some process which generates the measurements on a given unit.

Traditionally, randomization has been regarded as an essential part of survey sampling for objective inferences and estimability of the standard errors of estimates. In Case 1 above, randomization is created by the sampler through specific survey designs. This was the approach adopted by statisticians in developing the subject of sample survey. In this design based approach it is assumed that the population values $y_1,\ldots,y_N$ are fixed and hence $\underset{\sim}{y} = \{y_1,\ldots,y_N\}$ can be treated as a parameter of the population under consideration. Our interest is on some function of this parameter say $g(\underset{\sim}{y})$. There are certain authors,

for example, Neyman (1971) who tend to focus attention only on estimation based on man-made randomization in the form of design.

In Case 2, there are two sources of randomization, (i) created randomization based on survey design, and (ii) random error associated with the measurements of the sample unit which is commonly known as non-sampling error or bias. This latter aspect is beyond the scope of this thesis. Modern development of survey sampling techniques mainly follow the line of Case 3. This is widely known as the super-population approach or model based approach. Under this approach it is assumed that to each population unit is associated a random variable for which a stochastic structure is specified. The actual value associated with the population unit is treated as an outcome of that random variable. We shall discuss various super-population models and methods of estimation under these models in Chapter 2. Super-population approach is an elegant development of statisticians, through which important new methods are currently being added to traditional methodology of survey sampling. Authors like Barnard (1971), Kalbfeisch and Sprott (1969), Royall (1970, 1971), etc. consider inference based on super-population models not only desirable but almost necessary. Some of the authors have strongly criticized the idea of the sample design producing the only source of randomness in data injected by the survey statistician himself. "The survey statistician does not lean on probability-theory for the purpose of understanding and controlling the mess created by an unavoidable source of randomness or uncertainty (observation error)", Basu (1969). Basu examined the randomization principle in survey sampling and came to the conclusion that there is very little, if any, use for the survey designs. Chapter 3 deals with the randomization principle and its alternatives.

Although, we have called the common and well-known approach of survey sampling as traditional or conventional, the idea of super-populations is also not new. Cochran (1939, 1946), Deming and Stephan (1941), Madow and Madow (1944), Mahalanobis (1944) are early users of the super-population idea. Deming and Stephan (1941) were first to clearly mention the idea of variable status of the population, rather than fixed. They made the comment that the census is a sample only and suggested that it is one of many populations that might have resulted. The difference between census and sample survey is a matter of degree and considered the population census whose state of nature is changing with time. Cochran (1946) first clearly assumed that the finite population we have at our disposal is actually a sample from an infinite population. He considered the population in which the variance among the elements in any group of contiguous elements increases as the size of the group increases. This type of population was also considered by Smith (1938), Jessen (1942), Mahalanobis (1944) and Hansen and Hurwitz (1943). Various mathematical models have been considered by these authors for representing the situation where the variance within a group is directly proportional to the size of the group measure x_i. Cochran (1946) considered that elements x_i are drawn from different populations and assumed that the population changes in some regular manner with the value i. Alternatively, he suggested that x_i belongs to the same population but is serially correlated, and found it more reasonable to consider the finite population as a sample from an infinite population.

The idea of super-population models i.e. the idea of considering the existing population as a sample from an infite

population, started much earlier, unfortunately, the theoretical aspect of the model based approach did not attract much attention from statisticians until the 1960's. It is well-known that the use of supplementary information in estimation of finite population parameters, in general, increases the accuracy of the estimator. So, samplers felt the need for a comparison of the relative accuracy of sample designs using such information. This comparison becomes difficult if we cannot assume any functional relationship of the data. One solution to the problem is to regard the finite population as a random sample from an infinite super-population model having certain properties. The results so obtained do not apply to any single finite population but to the average of all finite populations that can be drawn from the infinite population, (Raj, 1958).

Early works on super-population models are based on some type of linear regression models with heteroscedastic error variances. Hacking (1965) has proposed the concept of "chance set-up" as a logical superior to the postulate of a hypothetical infinity of populations. For example, the linear regression super-population model may be viewed as defining a random set-up rather than an infinity of hypothetical populations, if so desired. But, the analyses are mathematically identical. Forman and Brewer (1971) have given comparisons of the efficiencies of six methods of sampling in common use. The model they used (also commonly used super-population model) is an infinite set of theoretical populations, each of size N. Units are identifiable, having two measures Y_i and X_i on the i^{th} unit, where Y_i is the measurement of the character of interest and X_i is measurement of the auxiliary information (e.g. size of the unit), and related to Y_i

as follows

$$(1.2) \qquad Y_i = \alpha + \beta X_i + e_i , \qquad i = 1,\ldots,N ,$$

where, α and β are constants, e_i's are random variables with $E(e_i) = 0$, $E(e_i^2) = \sigma_i^2$ (sometimes $E(e_i^2)$ is some function of X_i) and $E(e_i, e_j) = 0$, for all $i \neq j$. Here, expectation, E, is over all hypothetical populations and σ_i^2 is constant over all these populations but varies with i.

During the 1960's, statisticians have devoted much attention to the theoretical aspects of survey sampling. For a long time there were big gaps between survey sampling theories and statistical inference theories. In the traditional books of survey sampling, authors used the statistical inference theories under the assumption of large samples. The maximum likelihood method of estimation in statistical inference was essentially (for a long time) a failure in survey sampling situations. If the sample is drawn with probability proportional to size of the unit then how valid are traditional methods in the theory of hypothesis testing or the theory of statistical inferences? The answer of this question is still unknown. However, in the late 60's and early 70's it became possible to relate likelihood methods and Bayesian methods with finite populations. Some examples are Royall (1968, 1976a), Hartley and Rao (1968, 1969), Kalbfleish and Sprott (1970), C.R. Rao (1971), Ericson (1969a, b), Solomon and Zacks (1970), Basu (1969), Zacks (1969), and Godambe (1966, 1968), Godambe and Thompson (1971), Godambe and Joshi (1965), etc.

The most remarkable and striking development in survey sampling theory during the 1960's is the development of design free inferences.

There are some strong critics on the use of survey design for inference on finite populations. Godambe (1966) noted that the application of the likelihood principle in sampling situation would mean that the sampling design is irrelevant for data analysis. Basu (1969) examined the role of sufficiency and the likelihood principle and gave the conclusion, "Once the sample has been draw, the inference should not depend in any way on the sampling design. This poses the problem of designing a survey which will yield a good (representative) sample." He also examined the randomization principle (the man-made randomization through survey design) and pointed out very limited use, if any, for it in survey design.

Carrying this idea further, statisticians in the 1970's started suggesting the use of subjective sampling for an optimum estimator. Royall (1970) suggested a subjective sample, called a 'balanced sample', for estimating the population total. For estimating population total his estimator based on this balanced sample under the assumption of linear-regression super-population model proved to be most efficient. Brewer (1963) first suggested this type of purposive sampling. Later Royall (1973a, b) studied the robustness of the estimator based on balanced samples. This idea was further developed and extended by many other authors, namely, Holt (1975), Sigha (1976), Mukhopadhyay (1977), Tallis (1978), Scott, Brewer and Ho (1978), Singh and Garg (1979). There is considerable criticisms of this type of purposive sampling although the mathematical basis of this approach is sound. However, it seems that as yet there is no conclusive decision on the use of design based approach and model based approach.

Both approaches have some merits and demerits. Some authors are trying to mix these two streams. For example, Kolehmainen (1981) suggested that stratification of the finite population should always be made, if possible, and sampling within strata can be made purposively. Basu (1978) also suggested some type of post-stratification of data. In Chapter 3, we discuss this matter in greater detail.

In Chapter 4, we discuss the use of order statistics in the estimation of quantiles of finite populations. There we have given some results on properties of the distribution of order statistics in finite population sampling, bivariate distribution of sample quantiles and estimation of quantiles using auxiliary information. In Chapter 5, we discuss the asymptotic behavior of some estimators of finite population parameters and derive the asymptotic joint distribution of sample quantiles.

CHAPTER II

SUPER-POPULATION MODELS AND PREDICTION

§2.1 INTRODUCTION

In this chapter we shall study different types of super-population models and sampling theories based on these models. Ideas and write-up of this chapter are mostly as in Cassel, Sarndal and Wretman (1977).

In Chapter 1, we have mentioned that the super-population model arises when we consider the measurement $\underset{\sim}{y} = (y_1,\ldots,y_N)$ of a finite population to be the outcome of a random variable, $\underset{\sim}{Y} = (Y_1,\ldots,Y_N)$. Let us denote the joint distribution of $\underset{\sim}{Y}$ by ξ. Before proceeding to the next section let us introduce some useful definitions.

Ordered sample: A sequence $s^* = (k_1,\ldots,k_{n(s^*)})$ such that $k_i \in \mathcal{U}$ for $i = 1,\ldots,n(s^*)$ is called an ordered sample. The number of components of s^*, denoted by $n(s^*)$, is called the sample size.

□

Unordered sample: A non-empty set s such that $s \subseteq \mathcal{U}$ is called an unordered sample. The number of elements of s, denoted by $v(s)$, is called the effective sample size.

□

If the context is only with unordered sample, then we shall call unordered sample and effective sample size simply by sample and sample

size respectively. The set of all sets s will be denoted by $\mathcal{S}$.

Unordered sample design (or simply sample design): A function $p(s)$ on $\mathcal{S}$ satisfying $p(s) \geq 0$ for all $s \in \mathcal{S}$ and $\sum_{\mathcal{S}} p(s) = 1$ will be called an unordered sample design. Some authors refer to the pair $(\mathcal{S}, p(.))$ as the design.

□

The definition of the ordered sample design is similar.

Non-informative design: A sample design $p(.)$ is called a non-informative design, if and only if, $p(.)$ is a function that does not depend on the y-values associated with labels in s or s^*. But $p(.)$ may be function of auxiliary variables.

□

Fixed size design: If $n(s^*)$ or $v(s)$ are fixed then the respective design is called a fixed size design.

□

§2.2 DIFFERENT SUPER-POPULATION MODELS

By super-population model or simply "model" we shall refer to a class of distributions ξ with various types of specifications. These specifications may be only on the first few moments of ξ or to be more specific we may assume ξ has some specific well-defined statistical distribution. However, in both cases it is assumed that the vector of finite population values $\underline{y} = (y_1, \ldots, y_N)$ is an out-

come of the random variable $\underset{\sim}{Y} = (Y_1,\ldots,Y_N)$ having distribution ξ.

Definition: If $Q = Q(Y_1,\ldots,Y_N)$ is a function of $Y_1,\ldots,Y_N$, the ξ-expectation of Q, denoted by $\mathcal{E}(Q)$ is defined as

$$\mathcal{E}(Q) = \int Q d\xi \ , \tag{2.1}$$

and ξ-variance of Q, denoted by $\mathcal{V}(Q)$ is defined as

$$\mathcal{V}(Q) = \int [Q-\mathcal{E}(Q)]^2 d\xi \ . \tag{2.2}$$

If $Q_1 = Q_1(Y_1,\ldots,Y_N)$ and $Q_2 = Q_2(Y_1,\ldots,Y_N)$ are two functions of $Y_1,\ldots Y_N$, the ξ-covariance of Q_1 and Q_2, denoted by $\mathcal{C}(Q_1,Q_2)$, is defined as

$$\mathcal{C}(Q_1,Q_2) = \int \{Q_1 - \mathcal{E}(Q_1)\}\{Q_2 - \mathcal{E}(Q_2)\} d\xi \ . \tag{2.3}$$

In particular, we shall define for $k = 1,\ldots,N$,

$$\mu_k = \mathcal{E}(Y_k), \quad \sigma_k^2 = \mathcal{V}(Y_k), \quad \sigma_{k\ell} = \mathcal{C}(Y_k,Y_\ell) \quad \text{for } k \neq \ell \ ,$$

$$\bar{\mu} = \frac{1}{N}\sum_1^N \mu_k \quad \text{and} \quad \bar{Y} = \frac{1}{N}\sum_1^N Y_k \ . \tag{2.4}$$

□

There are two broad classifications of models used in survey sampling. They are (i) general models, denoted by G and (ii) exchangeability models, denoted by E. Often there will be subscripts to further specify both Model G and Model E.

Model G_T (transformation model). This model specifies the class of distributions ξ such that, for given $a_k > 0$ and b_k, the variables

$$Z_k = \frac{Y_k - b_k}{a_k} , \qquad k = 1,\ldots,N$$

have common mean μ, variance σ^2 and covariance $\rho\sigma^2$ for any pair $k \neq \ell$. Unless, otherwise stated, in general μ, σ^2 and ρ are unknown, $-\frac{1}{N-1} \leq \rho < 1$, and $\sum_1^N a_k = N$. The condition on ρ is required to have non-negative $\mathcal{V}(\bar{Y})$. Therefore, under Model G_T, Y_k has the following moments:

$$\begin{aligned} \mu_k &= \mathcal{E}(Y_k) = a_k\mu + b_k \\ \sigma_k^2 &= \mathcal{V}(Y_k) = a_k^2\sigma^2 \\ \sigma_{k\ell} &= \mathcal{C}(Y_k,Y_\ell) = a_k a_\ell \rho\sigma^2 , \qquad k \neq \ell . \end{aligned} \tag{2.5}$$

Model G_T implies that the first two moments of the transformed variables $Z_1,\ldots,Z_N$ are unchanged. So we can suitably choose a_k and b_k for specifying a good sample design for the problem in hand.

Model G_{To}. The special case of Model G_T where $a_k = 1$, $b_k = 0$ for all $k = 1,\ldots,N$ is Model G_{To}. This model expresses that labels are uninformative.

Model G_{MR} (multiple regression model). The class of distributions ξ such that $Y_1,\ldots,Y_N$ are independently distributed and

$$\mu_k = \mathcal{E}(Y_k) = \beta_0 + \sum_{i=1}^{q} \beta_i x_{ki}, \qquad \sigma_k^2 = \mathcal{V}(Y_k) = \sigma^2 u_k ,$$

where $\beta_0,\beta_1,\ldots,\beta_q$ and σ^2 are unknown and $x_{k1},\ldots,x_{kq}$, u_k is a set of known numbers for all k, $k = 1,\ldots,N$.

Model G_R (ratio model). The class of distributions ξ such that $Y_1,\ldots,Y_N$ are independently distributed, and

$$\mu_k = \mathcal{E}(Y_k) = \beta x_k, \quad \sigma_k^2 = \mathcal{V}(Y_k) = \sigma^2 u(x_k) , \quad k = 1,\ldots,N ,$$

where β and σ^2 are unknown. $u(.)$ is a known function and $x_1\ldots,x_N$ are known positive constants. A common assumption is $u(x_k) = x_k^g$, where g is known.

Next, let us consider various types of exchangeability models. In order to be exchangeable, the distribution ξ must be symmetric in accordance with the following definition.

Definition. Random variables $Y_1,\ldots,Y_N$ are called exchangeable if $Y_{r_1},\ldots,Y_{r_N}$ have, for every permutation $r_1,\ldots,r_N$ of $1,\ldots,N$, the same joint distribution, which is called an exchangeable distribution. □

The idea of exchangeable distribution in the context of finite population was given by Ericson (1965). Variables $Y_1,\ldots,Y_N$ themselves may be assumed to be exchangeable. However, it is usually assumed that the transformed Y_k, under change of origin and scale, are exchangeable.

Model E_T. This model defines the class of distributions ξ such that, for known $a_k > 0$ and b_k, $k = 1,\ldots,N$ satisfying $\sum_1^N a_k = N$, the random variables

$$Z_k = (Y_k - b_k)/a_k , \qquad k = 1,\ldots,N ,$$

have an exchangeable absolutely continuous distribution. Common mean,

variance and covariance implied by the exchangeability will be denoted by μ, σ^2 and $\rho\sigma^2$ respectively. The first and second order moments of Y_k are given by (2.5). Y_k's themselves become exchangeable in the following special case of Model E_T.

Model E_{T0}. The special case of Model E_T such that $a_k = 1$, $b_k = 0$, for all $k = 1,\ldots,N$.

Let us now consider the discrete exchangeable super-population model. This is mostly known as random permutation or random labeling models. This model was first used in Madow and Madow (1944) and not addressed again until Kempthorne (1969). Recent works on random permutation models are Royall (1970a), Ramakrishnan (1970), C.R. Rao (1971), Godambe and Thompson (1973), Rao (1975) and Rao and Bellhouse (1978). Under the random permutation model we assume that N population values of $\underset{\sim}{y}$ are fixed but are labeled at random. So that, each permutation $r = (r_1,\ldots,r_N)$ of $1,\ldots,N$ is assumed to have probability equal to $1/N!$ of being assigned as labels for the units. The equivalent statement is that the fixed but unknown number $y_1,\ldots,y_N$ are assigned randomly to units with labels $1,\ldots,N$ so that each permutation of $y_1,\ldots,y_N$ has a probability equal to $1/N!$ for fixed labels $1,\ldots,N$. Under both versions it is implied that there is no systematic relationship between labels and corresponding y_k values. This y-value corresponding to the k^{th} label can be regarded as the outcome of a random variable Y_k. Let us now consider the following general model.

Model E_{RP}. (random permutation model). The class of distributions ξ such that, for any fixed, unknown numbers $z_1,\ldots,z_N$ and for given numbers $a_k > 0$ and b_k, $k = 1,\ldots,N$ such that $\sum_1^N a_k = N$, the

random variables

$$Z_k = (Y_k - b_k)/a_k \quad , \qquad k = 1,\ldots,N$$

have an exchangeable distribution such that

$$P(Z_1 = z_{r_1},\ldots,Z_N = z_{r_N}) = 1/N!$$

for each permutation $r_1,\ldots,r_N$ of $1,\ldots,N$. The Model E_{RP} implies, for any n, $1 \le n \le N$, the marginal distribution

$$P(Z_{k_1} = z_{r_1},\ldots,Z_{k_n} = z_{r_n}) = 1/N^{(n)} \quad ,$$

for each $N^{(n)} = n!\binom{N}{n}$ different sequences $r_1,\ldots,r_n$ of n numbers chosen from $Z_1,\ldots,Z_N$, where the corresponding random variables are $Z_{k_1},\ldots,Z_{k_n}$ for a fixed subset of labels $k_1,\ldots,k_n$. The ξ-moments for the Z_k are for $k = 1,\ldots,N$,

$$\begin{aligned} &\mathcal{E}(Z_k) = \mu_z \\ &\mathcal{E}(Z_k-\mu_z)^2 = \sigma_z^2 \\ &\mathcal{E}(Z_k-\mu_z)(Z_\ell-\mu_z) = \frac{-\sigma_z^2}{N-1} \quad , \qquad k \neq \ell \quad , \end{aligned} \tag{2.6}$$

where the unknown μ_z and σ_z^2 are given by

$$\mu_z = \frac{1}{N}\sum_1^N Z_k \quad , \qquad \sigma_z^2 = \frac{1}{N}\sum_1^N (Z_k-\mu_k)^2 \quad .$$

Therefore (2.6) implies

$$\begin{aligned} \mu_k &= \mathcal{E}(Y_k) = a_k\mu_z + b_k \\ \sigma_k^2 &= \mathcal{E}(Y_k-\mu_k)^2 = a_k^2\sigma_z^2 \\ \sigma_{k\ell}^2 &= \mathcal{C}(Y_k,Y_\ell) = \frac{-a_k a_\ell \sigma_z^2}{N-1} \quad , \qquad k \neq \ell \quad . \end{aligned} \tag{2.7}$$

It is to be noted here that in general Y_k are not exchange-

able. But, in special cases when $a_k = 1$ and $b_k = 0$ for all $k = 1,\ldots,N$ so that $Z_k = Y_k$ in Model E_{RP}, then Y_k are exchangeable.

Model E_{RPo}. The special case of Model E_{RP} such that $a_k = 1$, $b_k = 0$, for $k = 1,\ldots,N$, $\mu_z = \mu_y = \frac{1}{N}\sum_1^N y_k$, and $\sigma_z^2 = \sigma_y^2 = \frac{1}{N}\sum_1^N (y_i-\mu_y)^2$. So under this model any one of $N!$ permutations of $y_1,\ldots,y_N$ is an equally likely outcome of the random variables $Y_1,\ldots,Y_N$.

Parametric Super-Population Models: Usually in parametric super-population models the joint distribution ξ of $\underset{\sim}{Y} = (Y_1,\ldots,Y_N)$ is assumed to have known shape but depends on the unknown parameter $\underset{\sim}{\theta} = (\theta_1,\ldots,\theta_r)$, $\underset{\sim}{\theta} \in \Theta$, the parameter space. Let us assume that the distribution ξ_θ is continuous and $g(y/\theta)$ is the density of $\underset{\sim}{Y}$.

Model G_{PI} (parametric independent). The class of absolutely continuous distributions ξ such that $Y_1,\ldots,Y_N$ are independently but not necessarily identically distributed. Their joint density function is:

$$g(y/\theta) = \prod_1^N g_k(y_k/\theta), \qquad \theta \in \Theta,$$

where $g_k(./\theta)$ is the density function of Y_k.

Model E_P (parametric). The class of absolutely continuous distributions ξ such that $Y_1,\ldots,Y_N$ are exchangeable; their joint density function is symmetric in its N arguments, being $g(\underset{\sim}{y}/\theta)$, $\theta \in \Theta$.

Model E_{PI}. The class of absolutely continuous distributions ξ such that $Y_1,\ldots,Y_N$ are independently and identically distributed (and hence

exchangeable), their joint density function being,

$$g(\underset{\sim}{y}/\theta) = \prod_{1}^{N} g(y_k/\theta), \qquad \theta \in \Theta ,$$

where $g(./\theta)$ is the density function for all Y_k.

§2.3 SOME DEFINITIONS AND TERMINOLOGIES

In survey sampling, we mostly deal with prediction of population total $y = \sum_{1}^{N} y_i$ or the population mean $\bar{y} = \frac{1}{N} \sum_{1}^{N} y_i$ and their standard errors. For inference on $\bar{y}$ we need data. Under the model based approach, we have in general two sources of randomization in data. We can represent the observed data by $d = \{(k,y_k);\ k \in s\}$, where $s \in \mathcal{S}$, the set of all unordered samples and $y_k \in R_1$ for all $k = 1,\ldots,N$. Data d is outcome of the random variable $\mathcal{D} = \{(k,Y_k);\ k \in S\}$, where S is random and for each realized value s of S, Y_k for $k \in s$ is random, also. At this stage we can define two more random variables, namely, $D_1 = \{(k,Y_k);\ k \in s\}$ and $D_2 = \{(k,y_k);\ k \in S\}$. In D_1, $S = s$ is fixed and in D_2, $Y_k = y_k$ is fixed for $k = 1,\ldots,N$. So we have the sample space of $\mathcal{D}$ taking value d:

$$\mathcal{X} = \{d : s \in \mathcal{S},\ \underset{\sim}{y} \in \Omega\} ,$$

where, usually $\Omega = R_N$, the N-dimensional Euclidean space.

Let us define the statistic $T = T(\mathcal{D})$. If $S = s$ is given then T depends on $Y_1,\ldots,Y_N$ only through Y_k, $k \in s$. On the other hand, if $Y_k = y_k$, $k = 1,\ldots,N$ is fixed or given, then T depends on only S, i.e. T depends on the design only. Now if we

use $T(\mathcal{D})$ for inference on $\bar{Y}$, then we shall call $T(\mathcal{D})$ a predictor or estimator of $\bar{Y}$. Hence we can replace $\mathcal{D}$ of the predictor $T(\mathcal{D})$ by D_1, to get a new predictor $T(D_1)$ for $\bar{Y}$. This is still a function of random variables Y_k, $k \in s$. We shall often use simply T for $T(\mathcal{D})$ or $T(D_1)$ just to indicate that T is a function of random variables Y_k, where k may be in S or in s respectively. On the other hand the random variable obtained from $T(\mathcal{D})$ for $Y_k = y_k$, $k = 1,\ldots,N$ is fixed but $k \in S$ will be written as $t(D_2)$. The value of $T(\mathcal{D})$ for $S = s$ and $Y_k = y_k$, $k \in s$ will be written as $t(d)$. Thus $t(d)$ is no longer a random variable and will be termed as the estimate or predicted value of $\bar{y}$. As above, we shall simply write t for $t(D_2)$ or $t(d)$. The small letter t will indicate that t is a function of the realized value y_k of Y_k for $k \in S$ or $k \in s$.

If T is a predictor of $\bar{Y}$ then, let us define p-expectation as,

$$E(T) = \sum p(s) \cdot T \ ,$$

p-variance as,

$$V(T) = \sum p(s)(T-E(T))^2$$

and p-mean-square-error (p-MSE) as,

$$MSE(T) = \sum p(s)(T-\bar{Y})^2 \ .$$

It is to be noted here that $E(T)$, $V(T)$ and $MSE(T)$ are functions of

random varaibles $Y_1,\ldots,Y_N$.

Definition. T is called a p-unbiased (design unbiased) predictor of $\bar{Y}$, if and only if, for a given design p, $E(t) = \bar{y}$ for all $\underline{y} = (y_1,\ldots,y_N) \in R_N$, where t is the realized value of T for $Y_k = y_k$, $k \in S$. The strategy (p,T) is called p-unbiased if T is a p-unbiased predictor under p.

□

Definition. T is called ξ-unbiased predictor of $\bar{Y}$, if and only if, for any distribution ξ, $\mathcal{E}(T-\bar{Y}) = 0$ for all $s \in \mathcal{S}$, where $\mathcal{E}$ is the expectation operator with respect to ξ.

□

Remark: A predictor can be p-unbiased but not ξ-unbiased and vice-versa. For example, if p is a simple random sampling plan, then under the model G_R, $T = \frac{N}{n} \sum_s Y_i$ is a p-unbiased predictor, but with $Y = \sum_1^N Y_i$,

$$\mathcal{E}\left(\frac{N}{n} \sum_s y_i - Y\right) = \beta\left(\frac{N}{n} \sum_s x_i - \sum_1^N x_i\right) \neq 0 \quad ,$$

and hence not ξ-unbiased. On the other hand the ratio predictor,

$$T = \left(\sum_s Y_i \Big/ \sum_s x_i \right) \sum_1^N x_i$$

is not p-unbiased but ξ-unbiased.

□

Definition. T is called pξ-unbiased predictor of $\bar{Y}$, if and only if, for given p and ξ, $\mathcal{E}E(T-\bar{Y}) = 0$.

Definition. If T_1 and T_2 are predictors such that for the given design p, $\mathcal{E}\,MSE(p,T_1) \leq \mathcal{E}\,MSE(p,T_2)$ for all $\xi \in \mathcal{C}$, a given class of super-populations, then T_1 is called at least as good a predictor as T_2 for the design p. If strict inequality holds for at least one $\xi \in \mathcal{C}$, then T_1 will be called better than T_2

□

Definition. If (p_1,T_1) and (p_2,T_2) are strategies such that $\mathcal{E}\,MSE(p_1,T_1) \leq \mathcal{E}\,MSE(p_2,T_2)$ for all $\xi \in \mathcal{C}$, then we shall say that (p_1,T_1) is at least as good a strategy as (p_2,T_2). If strict inequality holds for at least one $\xi \in \mathcal{C}$, then we say that (p_1,T_1) is better than (p_2,T_2)

□

If $t_1 = t_1(D_2)$ and $t_2 = t_2(D_2)$ are estimators of $\bar{y}$ and if $E(t_1-\bar{y})^2 \leq E(t_2-\bar{y})^2$ for all $\underset{\sim}{y} \in R_N$, then $\mathcal{E}E(T_1-\bar{Y})^2 \leq \mathcal{E}E(T_2-\bar{Y})^2$ for any super-population model ξ, where T_1 and T_2 are predictors of $\bar{Y}$ corresponding to estimators t_1 and t_2 respectively. Hence if t_1 is at least as good as t_2 for estimating $\bar{y}$, for a given design p, then T_1 is at least as good as predictor T_2 for any ξ.

Lemma 2.1 (Cassel et al., 1977). Let T be any predictor of $\bar{Y}$. For any ξ and for any non-informative design p,

$$(2.8) \qquad \mathcal{E}\,MSE(p,T) = E\,\mathcal{V}(T)+E[\mathcal{B}(T)]^2+\mathcal{V}(\bar{Y})-2\,\mathcal{E}[(\bar{Y}-\bar{\mu})E(T-\bar{\mu})] \ ,$$

where, $\mathcal{V}(T) = \mathcal{E}(T-\mathcal{E}(T))^2$ and $\mathcal{B}(T) = \mathcal{E}(T-\bar{Y})$ are ξ-variance and ξ-bias of T respectively. In particular:

(a) If T is p-unbiased then

(2.9) $$\mathcal{E}V(T) = E\mathcal{V}(T) + E[\mathcal{B}(T)]^2 - \mathcal{V}(\bar{Y})$$

(b) If T is p- as well as ξ-unbiased then

(2.10) $$\mathcal{E}V(T) = E\mathcal{V}(T) - \mathcal{V}(\bar{Y}) \quad .$$

□

In the next two sections we are going to discuss various predictors under design oriented super-population model and design-independent super-population model.

§2.4 PREDICTION UNDER DESIGN ORIENTED SUPER-POPULATION MODEL

Since the publication of the paper by Horvitz and Thompson (1952) the estimator T_{HT}, well known as Horvitz-Thompson estimator, is considered in traditional literature of survey sampling as the most attractive estimator. Though it possesses some good optimal properties, but after development of super-population ideas it has lost some of its attractiveness. The Horvitz-Thompson estimator is defined for any arbitrary design as,

(2.11) $$T_{HT} = \sum_S \frac{Y_k}{N\alpha_k} \quad ,$$

where α_k is the inclusion probability of unit $k = 1,\ldots,N$. Basu (1971) suggested a modified form of T_{HT} which is known as the generalized difference predictor and is defined as:

Given an arbitrary vector $\underset{\sim}{e} = (e_1,\ldots,e_N)$ and a design with inclusion probabilities $\alpha_k > 0$, $k = 1,\ldots,N$, the generalized difference predictor is given by

(2.12) $$T_{GD} = \sum_{S} \frac{Y_k - e_k}{N\alpha_k} + \bar{e} \quad ,$$

where $\bar{e} = \sum_{1}^{N} e_k / N.$

The estimator T_{GD} has the following properties:

(i) T_{GD} is p-unbiased;

(ii) T_{GD} has zero p-variance for any value $\underset{\sim}{y}$ of $\underset{\sim}{Y}$ that satisfies $(\underset{\sim}{y}-\underset{\sim}{e}) \propto \underset{\sim}{\alpha} = (\alpha_1,\ldots,\alpha_N)$, provided that p is a design with fixed effective sample size, n, abbreviated as FES(n);

(iii) T_{GD} is ξ-unbiased for any model if $e_k = \mu_k$, $k = 1,\ldots,N$;

(iv) T_{GD} reduces to T_{HT} (a) if $\underset{\sim}{e} = 0$ or (b) $e_k \propto \alpha_k$, $k = 1,\ldots,N$ and p is FES(n) design;

(v) T_{GD} is origin and scale invariant.

If p is an FES(n) design with $\alpha_k > 0$, $k = 1,\ldots,N$, then $\mathcal{E}V(p,T_{GD})$ is minimized for the choice $e_k = \mu_k$ and $\alpha_k = fa_k$, $k = 1,\ldots,N$, where $f = n/N$. So, the optimal strategy of type (p,T_{GD}) is given by (p_o,T_{GDo}), consisting of

(i) any FES(n) design,

(2.13) $$p_o = p_o(s) \quad , \quad \text{such that} \quad \alpha_k = fa_k \quad , \quad k = 1,\ldots,N \quad ;$$

(ii) the predictor,

(2.14) $$T_{GDo} = \sum_{S} \frac{Y_k - b_k}{Na_k} + \bar{b} \quad , \qquad \bar{b} = \frac{1}{N}\sum_{1}^{N} b_k \quad .$$

The predictor T_{GDo} has the following properties:

(i) T_{GDo} is ξ-unbiased and $p\xi$-unbiased for any ξ satisfying model G_T and for any FES(n) design p.

(ii) T_{GDo} is p-unbiased under $p = p_o$, but for any arbitrary FES(n) design T_{GDo} is not necessarily p-unbiased.

Let us now consider the predictor T such that $T \in \mathcal{L}_u$, the class of all p-unbiased linear predictors of $\bar{Y}$. Hence, $E(T) = \bar{Y}$ and T is of the form:

$$T = w_{oS} + \sum_S w_{kS} Y_k \quad . \tag{2.15}$$

Theorem 2.1 (Cassel et al. (1976). Under Model G_T, and for $A = \frac{1}{N} \sum_1^N a_k^2$,

$$\mathcal{E}V(p,T) \geq \mathcal{E}V(p_o,T_{GDo}) = \frac{(1-\rho)(1-fA)\sigma^2}{n} \quad , \tag{2.16}$$

for any strategy (p,T) such that p is an FES(n) design with $\alpha_k > 0$, $k = 1,\ldots,N$ and $T \in \mathcal{L}_u$; equality holds if and only if $(p,T) = (p_o,T_{GDo})$.

□

Remark. A strategy based on the p-unbiased predictor T_{HT} as in (2.11) can never, under model G_T, be better than the strategy (p_o,T_{GDo}).

□

In the light of above theorem, in general, under model G_T it is advisable for an optimal predictor to use a design giving large inclusion probabilities of units considered by the model to be highly

variable. However, if all Y_k are assumed to have equal variances, that is, $a_k = 1$, for all $k = 1,\ldots,N$, then (p_o, T_{Do}) is the best strategy, where $p_o = p_o(s)$ is such that $\alpha_k = f$, for all $k = 1,\ldots,N$, satisifed for example in the case of simple random sampling and

$$T_{Do} = \bar{Y}_S + \bar{b} - \bar{b}_S \quad , \tag{2.17}$$

the well known difference estimator, where $\bar{b} = \frac{1}{N} \sum_S b_k$ and

$$\mathcal{E}\, V(p_o, T_{Do}) = \frac{(1-\rho)(1-f)\sigma^2}{n} \quad . \tag{2.18}$$

Under stratified random sampling, we know that if we have optimum allocation, the number of samples to be selected from a stratum is proportional to the variance of that stratum. That is we select larger numbers of units from a stratum having large variance. So, stratified random sampling is also a technique of unequal probability sampling for units not in the same stratum. In Horvitz-Thompson strategy we also give larger selection probability to units for which the variance is large, but here the problem of optimization was attacked from different angles only.

We have mentioned in Theorem 2.1, that the strategy (p_o, T_{GDo}) is optimal in the class of p-unbiased linear estimators. But if we have an exchangeable model, then as mentioned in Theorem 2.2 below, (p_o, T_{GDo}) is also optimum in the wider class of p-unbiased estimators. We do not have to adhere to linear estimators only.

Lemma 2.2 (Cassel et al., 1977). Let p be any given FES(n) design with $\alpha_k > 0$, $k = 1,\ldots,N$. Then, under Model E_T,

$$(2.19) \qquad \mathcal{E}E(T-\bar{\mu})^2 \geq \mathcal{E}E(T_{GDo}-\bar{\mu})^2 = \frac{\{1+\rho(n-1)\}\sigma^2}{n} \quad ,$$

for any linear or non-linear $p\xi$-unbiased estimator T of $\bar{\mu}$; equality holds if and only if $T = T_{GDo}$.

Theorem 2.2 (Cassel et al., 1977). Under Model E_T, letting $A = \frac{1}{N}\sum_{1}^{N} a_k^2$,

$$(2.20) \qquad \mathcal{E}V(p,T) \geq \mathcal{E}V(p_o,T_{GDo}) = \frac{(1-\rho)(1-fA)\sigma^2}{n} \quad ,$$

for any strategy (p,T) such that p is an FES(n) design with $\alpha_k > 0$, $k = 1,\ldots,N$ and $T \in \mathcal{A}_u$, the class of all (linear or non-linear) p-unbiased predictors of $\bar{Y}$; equality holds if and only if $(p,T) = (p_o,T_{GDo})$, where p_o and T_{GDo} are given by (2.13) and (2.14) respectively. □

Theorem 2.2 is also true for random permutation model E_{RP}. An extensive investigation of random permutation models has been done by Rao and Belhouse (1978). Using generalized random permutation models and general class of linear estimators of finite population mean, Rao and Belhouse have shown that many of the conventional estimators are optimal in the sense of maximum average mean-square error. They investigated optimality under the following sample designs: unistage design, stratified design, post-stratified design, double sampling design, sampling on two occasions and two-stage sampling design.

§2.5 PREDICTION UNDER DESIGN FREE SUPER-POPULATION MODEL

Most of the survey statisticians, who believe in the model based approach of estimation in survey sampling argue that p-unbiased-

ness is an unnecessarily heavy restriction, and instead ξ-unbiasedness or possibly $p\xi$-unbiasedness should be required. Their opinion is, average of $(T-\bar{Y})^2$ with respect to design p is a matter of presampling interest only. In this section, we shall deal with the design-free model based approach of prediction. Here the distribution ξ is the essential element of inference, where s is treated as given, giving less attention to design p producing the sample s. So, our object is to choose T, for any given s, to minimize $\mathcal{E}(T-\bar{Y})^2$. The average with respect to p is of secondary importance. It turns out that the predictor T that minimizes $\mathcal{E}(T-\bar{Y})^2$ for any given s is also the predictor that minimizes $\mathcal{E}\,E(T-\bar{Y})^2$ for any given non-informative design p.

Here we shall assume that super-population distribution ξ_θ, depends on certain parameter (or parameter vector) $\theta \in \Theta$, which is unknown. Once we can specify ξ_θ, the method of prediction of $\bar{Y}$ becomes a classical inference problem.

For an arbitrary set $s \in \mathcal{S}$, let $\xi_s = \xi_{s,\theta}$ be the marginal distribution of $Y_{k_1},\ldots,Y_{k_{v(n)}}$ where $k_1 < \ldots < k_{v(s)}$ is an enumeration in increasing order of the labels $k \in s$ and let $\xi_{\bar{s}/s} = \xi_{\bar{s}/s,\theta}$ be the joint conditional distribution of Y_k, $k \in \bar{s}$, (taken in increasing order of k) given $Y_{k_1},\ldots,Y_{k_{v(s)}}$. Let the corresponding density function be $g(y/\theta)$, $g_s(y_s/\theta)$, $g_{\bar{s}/s}(y_s/\theta)$. Note that if ξ is an exchangeable distribution, then so are ξ_s and $\xi_{\bar{s}/s}$. Let $\mathcal{E}$, $\mathcal{E}_s$ and $\mathcal{E}_{\bar{s}/s}$ be expectation operators associated with ξ, ξ_s and $\xi_{\bar{s}/s}$ respectively. Now, if p is non-informative, which we shall in general assume, then the operators $\mathcal{E}$ and E may be interchanged, that is

$$(2.21) \qquad \mathcal{E}\mathrm{MSE}(p,T) = \mathcal{E}E(T-\bar{Y})^2 = E\,\mathcal{E}(T-\bar{Y})^2 = E\,\mathcal{E}_s\mathcal{E}_{\bar{s}/s}(T-\bar{Y})^2 \quad .$$

Here our objective is to minimize $\mathcal{E}(T-\bar{Y})^2$ for any sample s. So if we can find a T^* which minimizes $\mathcal{E}(T-\bar{Y})^2$ for any $s \in \mathcal{S}$, and p is any non-informative design, then T^* also has the property of minimizing $\mathcal{E}\mathrm{MSE}(p,T)$ for any given design p. Alternatively, if T^* is such that it minimizes $\mathcal{E}(T-\bar{Y})^2$, then in the presampling stage we can look for the best design p which uses T^* and minimizes $\mathcal{E}\mathrm{MSE}(p,T^*)$ for different p.

Let the population mean

$$(2.22) \qquad \bar{Y} = f_s\bar{Y}_s + (1-f_s)\bar{Y}_{\bar{s}} \quad ,$$

where, $f_s = \frac{v(s)}{N}$, $\bar{Y}_s = \frac{1}{v(s)}\sum_s Y_k$ and $\bar{Y}_{\bar{s}} = \frac{1}{N-v(s)}\sum_{\bar{s}} Y_k$. We realized the value $\bar{y}$ of $\bar{Y}$ and this can be expressed as

$$(2.23) \qquad \bar{y} = f_s\bar{y}_s + (1-f_s)\bar{y}_{\bar{s}} \quad .$$

In this representation of population mean $\bar{y}$, the first part of the right hand side is known due to sampling. So, Basu (1971) suggested that attempts should be made for a post survey estimation of the unknown part $\bar{y}_{\bar{s}}$. But this idea is criticized by the decision-theorist as here the estimator is selected after observing data.

Let U be a predictor of $\bar{Y}_{\bar{s}}$, then it follows from (2.22), for any given sample s,

$$(2.24) \qquad T = f_s\bar{y}_s + (1-f_s)U$$

is a predictor of $\bar{Y}$. Since s is given, the distribution $U = U(D_1)$

and $T(D_1) = f_s\bar{Y}_s + (1-f_s)U(D_1)$ depends entirely on ξ. In terms of U, the MSE can be written as

$$(2.25) \qquad \mathcal{E}\, E(T-\bar{Y})^2 = E\{(1-f_s)^2\, \mathcal{E}_S\, \mathcal{E}_{\bar{S}/S}\, (U-\bar{Y}_{\bar{S}})^2\} \quad .$$

If ξ is completely known, minimum $\mathcal{E}$MSE is obtained if, for any given s, we choose

$$(2.26) \qquad U = \mathcal{E}_{\bar{S}/S}\, (\bar{Y}_{\bar{S}}) \quad .$$

However, if ξ depends on the unknown parameter vector $\underset{\sim}{\theta}$, then at first we shall have to estimate $\underset{\sim}{\theta}$, and then attempt to predict $\bar{Y}$. Note that T is ξ-unbiased for $\bar{Y}$, if and only if, for every $s \in \mathcal{S}$, U is ξ-unbiased for $\bar{Y}_{\bar{S}}$.

Let us now consider some ξ-unbiased predictors. If we relax the condition of p-unbiasedness of the last section and impose the more loose restriction of ξ-unbiasedness then we can find predictors with smaller mean-square-errors. This is demonstrated by the following theorem of Cassel et al. (1977).

Theorem 2.3. Let p be any given design. Then, under Model G_T,

$$(2.27) \qquad \mathcal{E}\, E(T-\bar{Y})^2 \geq \mathcal{E}\, E(T^*-\bar{Y})^2$$

where T is any linear ξ-unbiased predictor of $\bar{Y}$ and for any $s \in \mathcal{S}$,

$$(2.28) \qquad T^* = f_s\bar{Y}_s + (1-f_s)(\bar{Z}_s\bar{a}_{\bar{s}} + \bar{b}_{\bar{s}})$$

where

$$(2.29) \qquad \bar{Z}_s = \frac{1}{v(s)} \sum_s Z_k \,, \qquad Z_k = \frac{(Y_k - b_k)}{a_k} \,,$$

$k = 1,\ldots,N$. Equality holds if and only if $T = T^*$. □

Next, let us consider the predictor T^* under various special cases of Model G_T.

(i) If under Model G_T, $b_k = 0$ for all $k = 1,\ldots,N$, then

$$T^* = \bar{Z}_s + \frac{1}{N}\sum_s (a_k - \bar{a}_s)Z_k \tag{2.30}$$

where $\bar{Z}_s$ is as in (2.29) and $Z_k = Y_k/a_k$. If p is an FES(n) design, then

$$T^* = T_{HTo} + \frac{1}{N}\sum_s (a_k - \bar{a}_s)Z_k \tag{2.31}$$

where $T_{HTo} = \sum_s Y_k/(na_k)$.

(ii) If under Model G_T, $a_k = 1$ for all k, then

$$T^* = \bar{Y}_s + \bar{b} - \bar{b}_s \tag{2.32}$$

the usual difference predictor.

(iii) Finally, under Model G_{To},

$$T^* = \bar{Y}_s$$

the sample mean.

For the sake of comparison let us discuss the intuitively appealing predictor,

$$T^o = \bar{Z}_s + \bar{b} \quad . \tag{2.33}$$

This predictor has the following properties.

(i) T^o is ξ-unbiased predictor of $\bar{Y}$ under Model G_T.

(ii) T^o minimizes, under Model G_T, for any fixed s, the criterion $\mathcal{E}(T - \bar{\mu})^2$, among linear ξ-unbiased estimators of the super-population parameter $\bar{\mu} = \mu + \bar{b}$; hence T^o would be preferred if

inference were directly to the super-population and not to the realization $y_1,\ldots,y_N$.

(iii) If p is an FES(n) design, $T^o = T_{GDo}$ given by (2.14).

(iv) If all $b_k = 0$ and p is an FES(n) design, then $T^o = T_{HTo}$.

From the above discussion, it comes out that T^* and T^o are both optimal, but by different criteria. If the criteria is min $\mathcal{E}$ MSE then T^* is optimal and better than T^o to an extent as shown in the following theorem (Cassel et al, 1977).

Theorem 2.4. Under Model G_T, for any design p,

$$(2.34) \qquad \mathcal{E}E(T^o-\bar{Y})^2 - \mathcal{E}E(T^*-\bar{Y})^2 = \frac{E\{\sum_S (a_k-\bar{a}_S)^2\}(1-\rho)\sigma^2}{N^2} \geq 0$$

where T^* and T^o are given in (2.28) and (2.33). Moreover,

$$(2.35) \qquad \mathcal{E}E(T^o-\bar{Y})^2 = [E\{\frac{1}{v(S)} - \frac{A}{N}\} + \frac{2E(A-\bar{a}_S)}{N}](1-\rho)\sigma^2$$

where $\bar{a}_S = \frac{1}{v(S)} \sum_S a_k$, and $A = \frac{1}{N} \sum_1^N a_k^2$. Strict inequality holds in (2.34) if $p(s) > 0$ for some s such that not all a_k for $k \in s$ are equal. □

The comparison given in Theorem 2.4 hold for any p but neither T^* nor T^o are necessarily p-unbiased. However, both are ξ-unbiased under Model G_T.

Models G_{PI}, E_{PI}, G_R, G_{MR} assume that $Y_1,\ldots,Y_N$ are independently distributed. Under the assumption of independence, the optimal ξ-unbiased predictor is given by the following theorem.

Theorem 2.5. (Cassel et al., 1977). Let p be any given design, let $T = f_s \bar{Y}_s + (1-f_s)U$ and $T' = f_s \bar{Y}_s + (1-f_s)U'$ be two ξ-unbiased predictors of $\bar{Y}$. Then if ξ is a product measure (as under Model G_{PI}, E_{PI}, G_R, G_{MR}) the inequality

$$\mathcal{E}\, E(T-\bar{Y})^2 \leq \mathcal{E}\, E(T'-\bar{Y})^2$$

holds if and only if, for any $s \in \mathcal{S}$ such that $p(s) > 0$, and

$$\mathcal{V}(U) = \mathcal{E}(U-\bar{\mu}_s)^2 \leq \mathcal{V}(U') = \mathcal{E}(U'-\bar{\mu}_s)^2 .$$

If for some s with $p(s) > 0$ the latter inequality is strict, then the former inequality is also strict. □

A similar type of result was derived by Fuller (1970). In view of these results, much of classical parametric estimation is relevant to finite population sampling. If we know something about the shape of the distribution it is possible to construct predictors of $\bar{Y}$ which are more efficient than sample mean $\bar{y}_s$. For example Fuller (1970) proposed simple predictors of $\bar{Y}$ when the tail of the distribution is well approximated by the tail of a Weibull distribution and Ringer, Jinkins and Hartley (1972) proposed a square root predictor for a positively skewed population.

Much of the literature on super-populations contains discussion on models G_R and G_{MR}. We have already mentioned in Chapter 1, that idea of super-population first came from analysis of ratio and regression estimators. The following theorem due to Brewer (1963) and Royall (1970b) is the most important result under Model G_R. The theorem is true for any design p and gives the best linear ξ-unbiased (ξ-BLU) predictor of $\bar{Y}$. Here the best

is in the sense of min $\mathcal{E}$MSE. We shall denote this ξ-BLU predictor by T_{BR}. It also comes out from this theorem that T_{BR} does not depend explicitly on design but its $\mathcal{E}$MSE does.

Theorem 2.6. Under Model G_R, and for known auxiliary variable measurements $x_k > 0$, $k = 1,\ldots,N$, the ξ-BLU predictor of $\bar{Y}$ is, for any design p, given by

$$T_{BR} = f_s \bar{y}_s + (1-f_s)\hat{\beta}\,\bar{x}_{\bar{s}} \tag{2.36}$$

where

$$\bar{x}_{\bar{s}} = \frac{1}{N-v(s)} \sum_{\bar{s}} x_k \,,$$

$$\hat{\beta} = \sum_s \frac{x_k y_k}{u(x_k)} \Big/ \sum_s \frac{x_k^2}{u(x_k)} \quad , \quad \text{and} \tag{2.37}$$

$$\mathcal{V}(Y_k) = \sigma^2 u(x_k) \,.$$

Furthermore,

$$\mathcal{E}\text{MSE}(p,T_{BR}) = \frac{E\{(\sum_{\bar{s}} x_k)^2 \mathcal{V}(\hat{\beta}) + \sigma^2 \sum_{\bar{s}} u(x_k)\}}{N^2} \,, \tag{2.38}$$

where, $\mathcal{V}(\hat{\beta}) = \sigma^2 / \sum_s (x_k^2/u(x_k))$. □

Special Cases: Let us denote T_{BR} by T_{BRg} if $u(x) = x^g$. IF $u(x) = x^g$, as assumed in many earlier literature of survey sampling, we have the following special cases.

(i) If $u(x) = x$, i.e., $g = 1$, then $T_{BR1} = T_R$, the classical ratio predictor

(2.39) $$T_R = \frac{\bar{x}\,\bar{Y}_s}{\bar{x}_s}$$

with

(2.40) $$\mathcal{E}\,\mathrm{MSE}(p,T_R) = \frac{\bar{x}\,E\{N\,\bar{x} / \sum_s x_k - 1\}\sigma^2}{N} \quad .$$

The predictor T_R is ξ-unbiased.

(ii) If $u(x) = x^2$, i.e., $g = 2$, then

(2.41) $$T_{BR2} = \bar{x}\,R_{Yx} + f_s(\bar{Y}_s - \bar{x}_s R_{Yx})$$

where

(2.42) $$R_{yx} = \frac{1}{v(s)} \sum_s Y_k/x_k \quad .$$

If p is an FES(n) design, i.e., $v(s) = n$ and assuming $\alpha_k = \frac{nx_k}{N\bar{x}}$ in the Horvitz-Thompson predictor T_{HT} of (2.11), then we have $T_{HT} = \bar{x}R_{Yx}$. Royall (1970b) has shown that if (i) p is any FES(n) design, (ii) $nx_k/N\bar{x} < 1$ for $k = 1,\ldots,N$ and (iii) $u(x)/x^2$ is a non-increasing function (usually $0 < g < 2$), then under Model G_R,

$$\mathcal{E}\,\mathrm{MSE}(p,\ \bar{x}R_{Yx}) \geq \mathcal{E}\,\mathrm{MSE}(p,T_{BR2})$$

or in the present form of T_{HT},

$$\mathcal{E}\,\mathrm{MSE}(p,T_{HT}) \geq \mathcal{E}\,\mathrm{MSE}(p,T_{BR2}) \quad .$$

It is clear from the above that $\mathcal{E}\,\mathrm{MSE}$ of the strategy (p,T_{BR}) depends on p through $E(.)$ in (2.38). A pre-sampling judgment may be required as to how p should be chosen such that $\mathcal{E}\,\mathrm{MSE}(p,T_{BR})$ is minimized. Under model ξ, after the sample has already been selected, the inference problem is simply the classic one

of predicting unobserved random variable $\bar{Y}_{\bar{s}}$ and the sample s should be one which permits a good predictor. This idea of Royall (1970b) has been criticized by some authors for adopting purposive samples. Expression (2.38) can be rewritten as follows:

$$(2.43)\qquad \mathcal{E}\,MSE(p,T_{BR}) = E\{\mathcal{E}(T_{BR}-\bar{Y})^2\} = \frac{1}{N^2}E[(\sum_{\bar{S}} x_k)^2\,\mathcal{E}(\hat{\beta}-\beta)^2+\sigma^2\sum_{\bar{S}} u(x_k)].$$

Now if our objective is to find a design p for which this is minimum then we have two options;

(i) to select a sample which will give a good estimate of the expected value of the mean of non-sampled units, i.e., to choose s so that

$$\frac{1}{N^2}\,(\sum_{\bar{s}} x_k)^2\;\mathcal{E}(\hat{\beta}-\beta)^2$$

is small, or

(ii) to observe those y-values which have greatest variances, so that only sum of the least variable values are to be predicted, i.e., to choose s such that $\sum_{\bar{s}} u(x_k)$ is small.

So it turns out that for wide class of variance functions, the optimum strategy is to use T_{BR} with a purposive sample s of FES(n) which contains the n largest x-values of the population.

Formally, let $\mathcal{S}_n = \{s : v(s) = n\}$ and s^* be the set of labels such that

$$(2.44)\qquad \max_{s\,\in\,\mathcal{S}_n}\sum_{s} x_k = \sum_{s^*} x_k\ ,$$

and let the design $p^* = p^*(s)$, such that

$$p^*(s) = \begin{cases} 1 & \text{if} \quad s = s^* \\ 0 & \text{if} \quad s \neq s^* \end{cases} \qquad (2.45)$$

Then the theorem of Royall (1970b) follows:

Theorem 2.7. Let p be any FES(n) design, and let p^* be defined by (2.45). If $u(x)$ is non-decreasing and $u(x)/x^2$ is non-increasing, then, under Model G_R,

$$\mathcal{E}\,MSE(p,T) \geq \mathcal{E}\,MSE(p^*,T_{BR}) \ ,$$

where T is any linear ξ-unbiased predictor of $\bar{Y}$, and T_{BR} is given by (2.36). □

Use of this type of extreme design is open to much criticism. J.N.K, Rao (1975) points out that there are, no doubt, situations in which the extreme design p^* can be highly efficient for prediction of one y-mean. But in most of the surveys, we also estimate mean values of other characters. In such situations extreme sample is not likely to work well if several means have to be estimated in the same survey. So, it is preferred and safe to use simple random sample in the case of multipurpose studies.

It is also obvious, from the above, that the result depends too much on the assumed model. If Model G_R is not true, that is, if $\mathcal{E}(Y_k) = \beta x_k^m$, $m \neq 1$ ($m = 1$ in case of G_R), then ξ-bias of T_R is

$$\mathcal{E}(T_R - \bar{Y}) = \bar{x}\beta\left[\left\{\left(\sum_s x_k^m\right) \Big/ \sum_s x_k\right\} - \left\{\left(\sum_1^N x_k^m\right) \Big/ \sum_1^N x_k\right\}\right] \ . \qquad (2.46)$$

Simple random sample is likely to give small bias in such cases, but

extreme design p^* is supposed to produce higher ξ-bias.

Results based on Model G_R can be easily extended for Models G_{MR}. Various results under Model G_{MR} have been given by Hartley and Sielken (1975), Royall (1976), Royall and Cumberland (1978a) and Tallis (1978). To present some of these results, let us introduce the following notations.

$$\mathcal{E}(Y_s) = X_s\beta \ , \quad \mathcal{E}(Y_{\bar{s}}) = X_{\bar{s}}\beta$$

$$\mathcal{V}(Y_s) = \sigma^2 V_s \ , \quad \mathcal{V}(Y_{\bar{s}}) = \sigma^2 V_{\bar{s}} \ , \quad \mathcal{C}(Y_s, \bar{Y}_{\bar{s}}) = 0$$

where. Y_s is $v(s)$-vector, i.e., vector of sampled y_k-values, $k \in s$, $Y_{\bar{s}}$ is $\{N-v(s)\}$-vector, having non-sampled Y-values as its components. Let us further assume that in both cases y_k's are enumerated in order of increasing k. $\beta' = (\beta_1,\ldots,\beta_q)$ is a vector of unknown parameters. Known matrix X_s and $Y_{\bar{s}}$ are of order $v(s)\times q$ and $\{N-v(s)\}\times q$ respectively. Let the row vector corresponding to unit k of X_s or $X_{\bar{s}}$ be denoted by $X_k' = (x_{k1},\ldots,x_{kq})$ where $x_{k1} = 1$, for all values of k. So we have $(q-1)$ auxiliary variables, $x_{k2},\ldots,x_{kq}$ measured on each unit k of the population. Diagonal matrices V_s and $V_{\bar{s}}$ are of order $v(s)\times v(s)$ and $\{N-v(s)\}\times\{N-v(s)\}$ respectively. The diagonal element of unit k is u_k, a known quantity. Hence $\mathcal{V}(Y_k) = \sigma^2 u_k$, where σ^2 is unknown.

Under Model G_{MR} following theorem due to Cassel et al., (1977) gives the ξ-BLU predictor.

Theorem 2.8. Under Model G_{MR}, and for known auxiliary measurements X_s and $X_{\bar{s}}$, the ξ-BLU predictor of $\bar{Y}$, for any design p, is:

(2.48) $$T_{BLU} = f_s \bar{y}_s + (1-f_s)\, \underline{m}_s' \hat{\beta}_{BLU} \ ,$$

where, $\underline{m}_s' = (m_{s1}, \ldots, m_{sq})$, and for $i = 1, \ldots, q$

$$m_{si} = \sum_{s} x_{ki} / (N-v(s)) \ .$$

Moreover,

(2.49) $$\hat{\beta}_{BLU} = (X_s' V_s^{-1} X_s)^{-1} (X_s' V_s^{-1} Y_s) \ ,$$

$\mathcal{E}\,MSE(p, T_{BLU})$ is equal to p-expectation of

(2.50) $$\mathcal{E}(T_{BLU} - \bar{Y})^2 = \frac{\sigma^2}{N^2} \sum_{\bar{s}} u_k + (1-f_s)^2 \{\underline{m}_s' (X_s' V_s^{-1} X_s)^{-1} \underline{m}_s\} \sigma^2 \ .$$

□

It is clear that under models G_R and G_{MR}, ξ-unbiased predictors are weighted least square estimators. They do not depend on any particular design. On this point Scott and Smith (1974) says,

> "The fact that the estimators do not depend on the design p(.) may worry some people, but it seems to us that when prior knowledge is so strong that it can be specified by model of the form (1) (simple linear regression model) then the relationships expressed in the model should override the sampling scheme for certain purposes."

Obviously, model based inference depends very much on the model assumed. So the natural question is, what will be the behavior of optimal predictors if the assumed model is not true or deviate slightly. This leads us to study the robustness of predictors.

§2.6 ROBUSTNESS IN MODEL BASED INFERENCE

In real life, it is not known which model is producing our actual population. So, whenever we have doubt on the assumed super-population model, the correctness of results established in the preceeding sections becomes questionable. Royall and Herson (1973a, b) first discussed this problem with a polynomial regression model:

$$Y_k = h(x_k) + e_k , \qquad k = 1,\ldots,N \tag{2.51}$$

where

$$h(x_k) = \sum_{j=o}^{J} \partial_j \beta_j x_k^j \tag{2.52}$$

and $\partial_j = 0$ or 1 depending on whether the term x^j is present in the model or not. Also e_k's are uncorrelated random errors with $E(e_k) = 0$ and $V(e_k) = \sigma^2 u(x_k)$, $k = 1,\ldots,N$. They denoted this model as $\xi(\partial_0,\ldots,\partial_J : u(x))$. In our present notation this is a special case of Model G_{MR}. In particular if $\partial_0 = 0$, $\partial_1 = 1$ and $\partial_2 = \ldots = \partial_J = 0$, then the above model reduces to $\xi(0,1 : u(x))$ which is Model G_R. Let us consider the following two cases:

Case 1. Misspecification of variance function under Model G_R.

The Brewer-Royall predictor, T_{BR}, as defined in (2.36) and (2.37) is ξ-BLU predictor of $\bar{Y}$ under Model G_R. The form of T_{BR} obviously depends on the specification of $u(x)$. In previous sections we also introduced Model G_{Rg} which is Model G_R with $u(x) = x^g$ and the corresponding ξ-BLU predictor is T_{BRg}. If the assumed model is G_{Rgo}, then T_{BRgo} is supposed to be optimal. But, if it so happens

that the true model is G_{Rg1}, $g_1 \neq g_0$ then in general T_{BRgo} is no longer most efficient, although it is still ξ-unbiased. In this case the prefered predictor is T_{BRg1}. For fixed values of g_0 and g_1, neither of which is necessarily true value of g, the following theorem due to Royall (1970b) gives us indication for preference for one or the other predictor.

Theorem 2.9. If $0 \leq g_0 \leq g_1$, then, for any FES(n) design p, and for any specification of the function $u(x)$ in Model G_R such that $u(x)/x^{g_0}$ is non-increasing,

$$\hat{\mathcal{E}}\,\mathrm{MSE}(p,T_{BRgo}) \leq \hat{\mathcal{E}}\,\mathrm{MSE}(p,T_{BRg1}) \quad . \tag{2.53}$$

For any function $u(x)$ such that $u(x)/x^{g_1}$ is non-decreasing, the inequality in (2.53) is reversed. For strict inequality in (2.53), it is sufficient that $p(s) > 0$ for some s such that $v(s) = n$, and $x_k \neq x_\ell$ for some $k \neq \ell$ in s. □

Case 2. Misspecification in polynomial regression model.

Theorem 2.10. (Royall and Herson, 1973a). Under the model $\xi(\partial_0,\partial_1,\ldots,\partial_J : u(x))$ and for known auxiliary variable measurements $x_k > 0$, $k = 1,\ldots,N$, the ξ-BLU predictor of $\bar{Y}$ is for any design p is given by

$$T = f_s \bar{y}_s + (1-f_s) \sum_{j=o}^{J} \partial_j \hat{\beta}_j \bar{m}_{sj} \tag{2.54}$$

where, for $j = 0,\ldots,J$,

(2.55) $$m_{sj} = \sum_{s} x_k^j/(N-v(s))$$

and $\hat{\beta}_j$'s are the least square estimates of β_j's under the model $\xi(\partial_0,\partial_1,\ldots,\partial_J : u(x))$. □

This theorem gives ξ-BLU predictor of $\bar{Y}$ in the situation assuming that β_j's are estimable.

Royall and Herson (1973a) analyzed robustness using model $\xi(0,1 : x)$ and corresponding predictor, $T_{BR1} = T_R = \bar{x}\,\bar{y}_s/\bar{x}_s$. By Theorem 2.9, this predictor is ξ-BLU and is the classical ratio estimator. If, however, the alternative model $\xi(\partial_0,\partial_1,\ldots,\partial_J : u(x))$ is true, then the preferred predictor, for any design, is given by Theorem 2.10. Moreover T_R is ξ-biased under model $\xi(\partial_0,\ldots,\partial_J : u(x)) \neq \xi(0,1 : x)$. The bias of T_R is

(2.56) $$\mathcal{E}(T_R-\bar{Y}) = \sum_{j=o}^{J} \partial_j\beta_j\, m_1\, \{(m_{sj}/m_{s1}) - (m_j/m_1)\}$$

where

$$m_{sj} = \frac{1}{v(s)} \sum_{k\in s} x_k^j , \qquad m_j = \frac{1}{N} \sum_{k=1}^{N} x_k^j \quad .$$

It is clear from (2.56) that ξ-bias is zero if

(2.57) $$m_{sj}/m_{s1} = m_j/m_1$$

for all j such that $\partial_j = 1$ in the model $\xi(\partial_0,\ldots,\partial_J : u(x))$. The idea of balanced sample comes in survey sampling, from this relation Royall and Herson (1973a) defined balanced sample as follows:

Balanced Sample: A balanced sample denoted by $s(J)$ is a sample

satisfying (2.57) for $j = 1,\ldots,J$; that is $s(J)$ is such that

(2.58) $$m_{sj} = m_j , \qquad j = 1,\ldots,J .$$

A sample s such that (2.58) holds for $j = j_o < J$ is said to be balanced on the j^{th} moment. A design p which selects, with probability one, a balanced sample will be called a balanced (sample) design and will be denoted by pb_J. □

It is difficult to get a sample which is balanced up to J^{th} order. Simple random sample usually gives an approximately balanced sample. We shall discuss methods of approximating balanced samples and their alternatives in Chapter 3.

Using a balanced sample we can eliminate the bias incurred by the ratio estimator T_R, if the actual model is $\xi(1,1 : u(x))$. But to have this property for the estimator, he shall have to compensate for efficiency. Assuming $FES(n)$ design, Royall and Herson (1973a) compare the balanced sampling strategy $R_{BAL} = (pb_1, T_R)$ to $R_{OPT} = (p^*, T_R)$ which has mimimum $\mathcal{E}$MSE under $\xi(0,1 : x)$, where p^* is the optimum design as given in (2.45). We have

$$\mathcal{E}\,\mathrm{MSE}(R_{OPT}) = \min_{s \in \mathcal{A}_n} (\bar{x}/\bar{x}_s)(1-f)\bar{x}\sigma^2/n ,$$

and

$$\mathcal{E}\,\mathrm{MSE}(R_{BAL}) = (1-f)\bar{x}\,\sigma^2/n ,$$

where

$$\mathcal{A}_n = \{s : v(s) = n\} .$$

Therefore, efficiency loss is the absolute value of

(2.59) $$\min_{s \in \mathcal{S}_n} (\bar{x}/\bar{x}_s) - 1 \leq 0 \quad .$$

Royall and Herson (1973a) have given some numerical results on efficiency for different types of populations. General conclusion of their study is: Shape of the distribution is less important factor in determining the efficiency than is γ, the ratio of extremes of the distribution with finite lower and upper limits of range. Another result is that the protection against ξ-bias is often costly from an efficiency point of view.

The balanced design, pb_1, protects the predictor T_R against ξ-bias which would be incurred if $\xi(1,1 : u(x))$ not $\xi(0,1 : x)$ is the true model. The most attractive property of balanced sample design is that if conditions $m_{sj} = m_j$, $j = 1,\ldots,J$ are satisfied, then T_R is protected against ξ-bias under any model $\xi(\partial_0,\ldots,\partial_J : u(x))$. There is no additional loss of efficiency of (pb_J, T_R) relative to R_{OPT}. It is also observed that T_R reduces to $\bar{Y}_s$ under balanced sampling. Royall and Herson (1973a) have also shown that if $T = T(\partial_0,\ldots,\partial_J : u(x))$ denotes ξ-BLU predictor given by Theorem 2.10, then under balanced design pb_J,

$$T(1,\partial_1,\partial_2,\ldots,\partial_J : 1) = T(\partial_0,1,\partial_2, \ldots,\partial_J : x) = \ldots$$

$$\ldots = T(\partial_0,\partial_1,\ldots,\partial_{J-1},1 : x^J) = \bar{Y}_s \quad ,$$

for any configuration $\partial_0,\partial_1,\ldots,\partial_J$ of 0 and 1's.

The idea of balanced sampling is also extended for the classical regression predictor,

(2.60) $$T_{REG} = \bar{Y}_s + \hat{\beta}(\bar{x} - \bar{x}_s) \quad ,$$

with

$$\hat{\beta} = \{ \sum_s (x_k - \bar{x}) Y_k \} / \{ \sum_s (x_k - \bar{x})^2 \} ,$$

which is ξ-BLU for model $\xi(1,1 : x)$. Now, if alternative model $\xi(\partial_0, \partial_1, \ldots, \partial_J : u(x))$ were actually true, then in general T_{REG} is ξ-biased. This bias can be removed if we choose the balanced design pb_J. For any balanced sample, the predictor T_{REG} also reduces to sample mean $\bar{Y}_s$.

Balancing the design is on the average equivalent to p-unbiasedness, and the prediction sought by balancing eliminates the efficiency gain realized under model based approach if we were willing to accept an extreme, purposive sample as a basis for inference.

Recently attempts have been made by many authors to compare and if possible to mix both design based approach and model based approach in survey sampling. Royall (1976a) and Scott and Smith (1969) have applied super-population model to two-stage sampling. Scott and smith derived results by using Bayesian techniques and established optimality among linear unbiased estimators. Royall (1976b) has studied linear least square prediction approach in two-stage sampling and then used a probability model to analyze various conventional estimators and certain estimators suggested by theory as an alternative to the conventional estimators. Sarndal (1978) has compared two approaches for estimating population mean. He showed that several of the conventional results can be obtained and reinterpreted through model based theory and found that the model based framework often offers advantages over the design based one when it comes to present a lucid

argument in favour of some given sampling procedure. Thompsen (1978) has given some examples where super-population ideas in survey sampling were applied to different surveys in Norway. Empirical studies of prediction theory has been done most recently by Royall and Cumberland (1978b, 1981).

CHAPTER III

RANDOMIZATION AND BALANCED SAMPLING

§3.1 RANDOMIZATION

Randomization is a well-known and widely used method of survey data collection and analysis. The main purpose of this method is to make objective inference and presenting results of survey in a convincing way to users. Keeping these and other advantages in mind, under varied population structures, survey statisticians have developed, in past years, different survey designs and hence various estimators for parameters of interest. These design based inferences are still overwhelmingly in use. Since 1960, design based inference has faced new challenges. This has been briefly discussed in Chapter 1.

The method of maximum likelihood is still one of the most important ways of estimation in statistics. However, for a long time, likelihood method was essentially a failure in survey sampling especially under design based approach. For any design $p(.)$ and for any population vector $\underset{\sim}{y} = (y_1,\ldots,y_N)$, treated as a parameter, the probability that the random quantity D_2 will take a value $d = \{(k,y_k) : k \in s\}$ is given by

$$(3.1) \qquad p_y(d) = \begin{cases} p(s) & \text{if } d \text{ is consistent with } y, \\ & \text{or if } y \in \Omega_d \\ 0 & \text{otherwise} , \end{cases}$$

where a specified value $d = \{(k,y_k),\ k \in s\}$ is said to be consistent

with a population vector $y_0 = (y_{01},\ldots,y_{0N})$, if and only if, $y_k = y_{0k}$ for all $k \in s = \{k_1,\ldots,k_n\}$, the sample. Ω_d is the set of all $\underset{\sim}{y} \in R_N$ such that d is consistent with $\underset{\sim}{y}$.

It follows from (3.1) that $Pr(D_2 = d/S = s) = 1$ if d is consistent with $\underset{\sim}{y}$, and zero otherwise. If our interest is for the parameter $\underset{\sim}{y}$ and if the design is uninformative, then from (3.1) we also find that the likelihood function $L(\underset{\sim}{y}/d) = p_{\underset{\sim}{y}}(d)$ is independent of $\underset{\sim}{y}$. That is likelihood is flat, so every consistent value of $\underset{\sim}{y}$ is equally likely and no unique maximum likelihood estimator is available. Likelihood function of the form (3.1) which is not informative in nature was first studied by Godambe (1966). But with a super-population model at the back of the finite population, the appropriate likelihood function may be more informative, Royall (1976a). In view of (3.1), when the likelihood principle is applied to the survey sampling under fixed population approach has the following two consequences.

(i) Inference from survey data should be independent of the sample design.

(ii) The only inference about $\underset{\sim}{y}$ sanctioned by likelihood principle is the trivial one that the components y_k for $k \in s$ must coincide with the observed values. It does not admit discrimination among the possible values of the unobserved components of $\underset{\sim}{y}$, since all the values of $\underset{\sim}{y} \in \Omega_d$ have the same likelihood.

However, with a somewhat different point of view another likelihood function emerges which can yield a maximum likelihood estimate of $\underset{\sim}{y}$ under certain conditions, Royall (1968) and Hartly and Rao (1968, 1969).

Very interesting and detailed discussions on likelihood function, sufficiency, randomization, etc. for finite population sampling are given in a series of papers, namely Basu (1969, 1971, 1978). Basu wrote in the summary of his 1969 paper, "We examine the role of sufficiency and likelihood principle in the analysis of survey data and arrived at the revolutionary but reasonable conclusion that, once the sample has been drawn, the inference should not depend in any way on the sampling design. This poses the problem of designing a survey which will yield a good (representative) sample. The randomization principle is examined from this view point and it is noticed that there is very little, if any, use for it in survey design." In this design based approach of sampling theory, as we mentioned earlier, there is only one source of randomization in the data. The artificial randomization created by the sampler himself is not inherent to the problem. All results of conventional theory are based on this randomization. Basu (1969) suggests in the Bayesian point of view: Once the data d is in our hand, forget about the sampling plan $(\mathcal{S}, p(.))$, which is an artificial source of randomization. In the Bayesian plan for selecting the data d, there is no place for symmetric dice or random number tables. But, unfortunately, until recently sufficient attention has not been given to the problem. Basu suggests that any reasonable Bayesian sampling strategy would have the following characteristics.

(a) The sampling plan would usually be sequential. The statistician should continue sampling (one or a few units at a time) until he is satisifed with the information thus obtained or until he reaches the end of his resources (time and cost). His decision to select the units for

a particular sampling stage would depend (non-randomly) on the sample obtained in the previous stages.

(b) The probability that the statistician would end up observing the units $s = (k_1,k_2,\ldots,k_n)$ in this order, would depend on s and the state of nature $\underset{\sim}{y}$. This probability would be degenerate i.e., zero for some values of $\underset{\sim}{y}$ and unity for the rest of the values of $\underset{\sim}{y}$.

We have already mentioned that viewing the likelihood function from a different angle some authors arrived at different types of likelihood which readily yield a unique maximum. But it is to be noted here that in all those likelihood functions they ignored the label part k of the data $d = \{(k,y_k) : k \in s\}$ and considered the unlabeled data $d_s = \{y_k : k \in s\}$. Basu (1971) pointed out that the label past k is an ancillary statistic, that is, sampling distribution of the statistic k does not involve the state of nature $\underset{\sim}{y} = (y_1,\ldots,y_N)$. The sampling distribution of $k = (k_1,\ldots,k_n)$ is uniquely determined by the sample design. It is therefore obvious that the label part of the data cannot, by itself, provide any information about y. Knowing k, we only know the names (labels) of the population units that are selected for observation. Usually, we incorporate the prior knowledge of the auxiliary variable $x = (x_1,\ldots,x_N)$ in the sampling plan. But this does not alter the above situation. The label, k of the data d will still be an ancillary statistic. Now the question is: If the label part k is informationless then, does the observation part of the data, namely, d_s, contain all the available information about $\underset{\sim}{y}$? Basu (1971) answers this question with a definite "no" and says that a great deal of the information will

be lost if the label part of the data is suppressed. Without the knowledge of k, the surveyor cannot relate the components of the observation vector $\underset{\sim}{y}$ to the population units and so he cannot make any use of the auxiliary character $\underset{\sim}{x} = (x_1,\ldots,x_N)$ and whatever other prior knowledge he may have about the relationship between y and x.

Basu (1978), has given a counter example (Example 4.1) where the optimum sampling plan would be sequential and non-randomized. In that example it is very difficult to get any justification for random sampling. Randomization is deeply rooted in statistics which is quite difficult to ignore with some counter examples. In view of the above mentioned statisticians, the main use of randomization is to safeguard the sample against unknown biases. Like the conventional approach, survey design can no more be the only determinant for judging the quality of the data. Basu (1978) suggests that the principal determinant of how a particular datum ought to be analyzed is the datum itself. The key concept in survey theory ought to be the notion of poststratification. "Randomization is widely recognized as a basic principle of statistical experimentation. Yet we find no satisfactory answer to the question, Why randomize?", Basu (1980).

§3.2 BALANCED SAMPLING

Prsently it is a general feeling of statisticians that artificial randomization in survey sampling should not be the only means of inference. Purposive sampling is now-a-days increasingly getting justification for the analysis of survey data.

Purposive samples are subjective but there are some rationale and objective justifications available for them. From the discussion of the previous section, it is indicated that we need an alternative for the randomization principle. Prediction approach or Bayesian approach in survey sampling may work as useful alternatives. These approaches usually lead us to the selection of purposive samples. We have already mentioned in Chapter 2 that under certain super-population models, optimum sampling strategy (2.45) or balanced samples are more desirable than the random samples.

We have defined the balanced samaple in Chapter 2 and have discussed the robustness of the estimators under balanced samples. In this section we shall discuss how to get balanced samples and extensions.

Approximate Balanced Samples

Selection of exact balanced samples of higher order is a big practical problem. Usually, not all values of auxiliary variable x are known to the sampler. In such a case it is impossible to get a balanced sample. Even if all values of x are known, exact satisfaction of $\bar{x}_s^{(j)} = \bar{x}^{(j)}$, $j = 1,\ldots,J$, is usually impossible. However, when J and the sampling fraction f are small, it is easier to get approximate balanced sample $s(J)$. It is expected that a random selection of units is supposed to give an approximately balanced sample. The average value of $\bar{x}_s^{(j)}$, over all $\binom{N}{n}$ samples s, is $\bar{x}^{(j)}$ for $j = 1,2,\ldots$. So we can expect that random sample s is approximately $s(J)$. Simple random sample is supposed to yield fair approximation to $s(J)$ for $J \geq 1$, so when we use this approximate

balanced sample the ratio and regression estimators are approximately unbiased under J^{th} degree regression models. The estimators will also be approximately optimal under models where the variance of Y is a polynomial in x of degree J or less. It is true that there is possibility of large deviations of the sample from the balanced sample depending on the dispersion of x values. If this circumstance arises, then it is advisable to use restricted randomization, censoring or post-stratification in the data. Royall and Herson (1973a) have given an expression for the extent of bias in ratio estimators using approximate balanced samples.

In surveys ususally more than one auxiliary variable is available. It is very difficult to get a balanced sample with respect to all those characters. However, random selections will at least justify some degree of confidence that the selected sample is approximately representative. It is to be mentioned here that neither purposive selection of a balanced sample nor restricted randomization nor unrestricted random selection will guarantee balanced on other variables not explicitly considered in choosing the sample.

Extensions and Other Types of Balanced Samples

Up to this point we have only discussed the balanced sample suggested by Royall and Herson (1973a). There are other forms of balanced samples whose definitions depend on the super-population model under consideration. Some of these are direct extensions of the already defined balanced sample. Holt (1975) has extended the idea of balanced sample for a linear multiple regression model and defined the balanced

sample as the sample for which the first moment of each of p-auxiliary variables for the sampled and non-sampled part of the finite population are equal. Using this balanced sample he obtained BLU estimator for finite population total.

Scott, Brewer and Ho (1978) proposed an alternative to balanced sample which they called "overbalanced sample". Their overbalanced sample provides more efficient estimators than the balanced sample. Principal results of this article followed by using the model $\xi(0,1:V(x))$. Regardless of manner in which sample observations have been obtained, the BLU predictor of the population total Y under this model is

$$(3.2) \qquad T_0 = T_0[0,1:V(x)] = \sum_s Y_i + \left(\frac{\sum_s Y_i x_i / V(x_i)}{\sum_s x_i^2 / V(x_i)} \right) \sum_{\bar{s}} x_i \quad .$$

Let $s^*(J)$ be a particular sample for which,

$$(3.3) \qquad \frac{\sum_{\bar{s}} x_i^{(j)}}{\sum_{\bar{s}} x_i} = \frac{\sum_s x_i^{j+1}/V(x_i)}{\sum_s x_i^2/V(x_i)} \quad , \qquad j = 0,1,\ldots,J \quad ,$$

then the following Lemma and Theorem follows, (Scott, et al. (1978)).

Lemma 3.1. If $s = s^*(J)$, then T_0 is ξ-unbiased under the model $(\partial_0,\partial_1,\ldots,\partial_J : V^*(x))$ for any $V^*(x)$.

In fact, T_0 is the BLU predictor when $s = s^*(J)$ for a wide class of models.

Theorem 3.1. Suppose $s = s^*(J)$, then T_0 is the BLU predictor under the model $\xi(\partial_0,\partial_1,\ldots,\partial_J : V^*(x))$ for any variance function of the form

$$V^*(x) = V(x) \sum_{\partial=o}^{J} \partial_j a_j x^{j-1} \quad .$$

□

Special cases:

If $V(x) = x$ then we get the Royall and Herson (1973a) results leading the balanced samples, so that T_0 reduces to ordinary ratio estimator,

$$T_1 = \left(\sum_s Y_i \Big/ \sum_s x_i \right) \sum_1^N x_i \quad . \tag{3.4}$$

On the other hand if $V(x) = x^2$, T_0 becomes

$$T_2 = \sum_s Y_i + \left(\frac{1}{n} \sum_s Y_i/x_i\right) \sum_{\bar{s}} x_i \tag{3.5}$$

and (3.3) becomes

$$\sum_s \frac{x_i^{j-1}}{n} = \sum_{\bar{s}} x_i^j \Big/ \sum_{\bar{s}} x_i \,, \qquad j = 0,1,\ldots,J \quad . \tag{3.6}$$

Obviously, this is always true for $j = 1$. Scott et al. (1978) called samples satisfying condition (3.6) as "overbalanced". The mean square error of T_2 under $\xi(0,1:x^2)$ is

$$\sigma^2 \left\{ \sum_{\bar{s}} x_i^2 + \frac{1}{n} \left(\sum_{\bar{s}} x_i \right)^2 \right\} \quad . \tag{3.7}$$

If the sampling fraction is small and no single x_i dominates the

others, the MSE is affected very little by the choice of sample and little efficiency is lost by choosing an overbalanced sample.

In many practical situations, $V(x)$ increases more quickly than x but less quickly than x^2, so that $\xi(\beta_0,\beta_1,\ldots,\beta_J : V(x))$ with $V(x) = a_1^2 x + a_2^2 x^2$ is often a fairly realistic model. Both T_1 with balanced sample and T_2 with overbalanced sample are BLU for their respective samples under this model and it is interesting to compare their performances. Scott et al. (1978) shows that MSE of T_1 with balanced sample is

$$M_1 = N(N-n)(a_1^2\bar{x} + a_2^2\bar{x}^{(2)})/n \quad , \tag{3.8}$$

where $\bar{x}^{(2)} = \frac{1}{N}\sum_1^N x_i^2$, while MSE for T_2 with overbalanced sample is

$$M_2 = N(N-n)\bar{x}_{\bar{s}}(a_1^2 + a_2^2\bar{x})/n \quad , \tag{3.9}$$

where $\bar{x}_{\bar{s}}$ is the mean of x-values not included in the overbalanced sample.

It follows from (3.6) that if $j = 0$, then $\bar{x}_{\bar{s}} < \bar{x}_s$. So that, $M_1 > M_2$. Thus ratio estimator T_1 with balanced sample will be less efficient than using T_2 with overbalanced sample. The loss of efficiency will be small in general, if a_1^2 dominates a_2^2 but can be substantial if a_2^2 is relatively large. These results apply to any polynomial model $\xi(\beta_0,\beta_1,\ldots,\beta_J : V(x))$ with variance function $V(x) = a_1^2 x + a_2^2 x^2$; hence T_2 with overbalanced sample $(j = 0,1,\ldots,J)$ is more efficient than T_1 with balanced sampling of the same order.

How to get overbalanced sample:

If the sample selection is with probability proportional to x_i, then

$$E\Big(\sum_s x_i^{j-1}/n\Big) = \sum_1^N x_i^j \Big/ \sum_1^N x_i \quad . \tag{3.10}$$

This indicates that selection with probability proportional to x_i yields an approximate overbalanced sample if the sample size is large and sampling franction is small. On the other hand if the sampling fraction is large, then Scott et al. (1978) suggest selecting units with probability equal to

$$\frac{\lambda x_i}{1 + \lambda x_i} \tag{3.11}$$

where λ is the solution to the equation

$$\sum_i^N \frac{\lambda x_i}{1 + \lambda x_i} = n \quad . \tag{3.12}$$

Iterative solution for (3.12) is suggested with starting value $\lambda^o = n/\sum_1^N x_i$. Using (3.11), the probability of not sampling the i^{th} element is $(1 + \lambda x_i)^{-1}$, so that, for all j,

$$E\Big(\sum_s x_i^j\Big) = \sum_1^N \frac{x_i^j}{1 + \lambda x_i} = E\Big(\sum_s x_i^{j-1}\Big)/\lambda \quad . \tag{3.13}$$

It follows that,

$$E\Big(\sum_s x_i^{j-1}/n\Big) = E\Big(\sum_s x_i^j\Big) \,/\, E\Big(\sum_s x_i\Big) \quad , \tag{3.14}$$

such that, we would expect to obtain an approximately overbalanced sample, if the sample size is large enough.

Royall and Herson (1973b) and Scott et al. (1978) have extended their definition of balanced sampling and overbalanced sampling schemes respectively for stratified population. Royall (1976b) has given some different ideas of balanced sample for two-stage sampling. Let the finite population consists of N elements and K cluster with M_i elements in the i^{th} cluster, such that $\sum_{1}^{K} M_i = N$. Suppose, first we have chosen a sample s of k clusters and then from the sample i^{th} cluster a random sample s_i consisting of m_i elements has been selected out of M_i elements.

Let,

$$n = \sum_{s} m_i \ ,$$

$$\bar{M}^{(j)} = \sum_{i=1}^{K} M_i^j \Big/ K$$

$$\bar{M}_s^{(j)} = \sum_{s} M_i^j \Big/ n \ ,$$

then Royal (1976b) says that the above two-stage sample is balanced if,

(3.15) $$\bar{M}_s = \bar{M} \quad \text{and} \quad \bar{M}_s^{(2)} = \bar{M}^{(2)} \ .$$

This type of balanced sample gives unbiased ratio estimtors in two-stage sampling under a quadratic regression model and the ratio type estimator is best. This result also holds for higher order-polynomial models when the sample is balanced on the corresponding higher-order moments.

In the literature on survey sampling we also find a completely different definition of balanced sample given by Singh and Garg (1979). This is actually some kind of systematic sample with random start. The suggested balanced sample is: Assuming population size N and sample size n both even (for odd values N and n modification of the definition is also available), first draw $n/2$ units at random from the first $N/2$ units of the population and rest of the $n/2$ elements are taken from $(N/2+1)^{th}$ to N^{th} units of the population with indices $N+1-r_i$, $i = 1,2,\ldots,n/2$, where r_i is the index of the i^{th} unit drawn from first $N/2$ units.

This sampling plan has the advantage of both simple random sampling and systematic sampling and works best for population exhibiting linear trend or periodicity. Their empirical study shows that this balanced sampling is generally better than simple random sampling and in most of the cases even better than systematic sampling and stratified sampling.

It is clear from the above discussion of randomization and purposive sampling that there are some cases where the randomization principle does not carry much meaning but purposive selection like optimum sampling or balanced sampling etc., gives meaningful and higher percision estimators. It may be sometimes feasible to draw these type of purpositive samples, but in large scale surveys with many items the purposive design could lead to very inefficient estimators for some of the items. Rao (1975) says, "Of course, this criticism also applies to conventional designs such as the probability proportional to size sampling plans or stratification by size with a 100% sampling

rate in the stratum containing the units with largest x_i. In such a situation, it might be advisable to employ equal probability sampling and utilize any quantitative concomitant information only at the estimation state." The role of randomization in survey sampling cannot be taken as the only basis of data analysis and inference as the conventional survey samplers used to think.

CHAPTER IV

RANKS AND ORDER STATISTICS FOR FINITE POPULATION

§4.1 ORDER STATISTICS IN SAMPLING FROM FINITE POPULATION

There are few works in survey sampling literature on the use of ranks and order statistics in estimating finite population parameters. It seems that Wilks (1962, p. 243) is the first to discuss distribution of order statistics in samples from a finite population. There he considered a finite population π_N consists of N distinct elements, say $y_{01}, < y_{02} < \ldots < y_{0N}$, and derived the probability function of the sample r^{th} order statistics. Let s be a random sample of size n from this population and let us denote the order sample by $Y_{(1)} < Y_{(2)} < \ldots < Y_{(n)}$. So it follows that probability of the k^{th} order statistic of the sample being equal to the t^{th} order statistic of the population is

$$(4.1) \qquad P[Y_{(k)} = y_{0(t)}] = \frac{\binom{t-1}{k-1}\binom{N-t}{n-k}}{\binom{N}{n}} = p_{N,n,k}(t)$$

where $t = k,k+1,\ldots,N-n+k$.

We can consider $p_{N,n,k}(t)$ either as (i) the probability function of the random variable $Y_{(k)}$, its mass points being $y_{0(t)}$, $t = k,k+1,\ldots,N-n+k$, or (ii) the probability function of the random variable t, that is, the rank of the y-value in the population to which the k^{th} order statistic in the sample is equal.

Wilks (1962) has given the following results on moments of t. Moments of t are easier to derive from the following relation

$$(4.2) \qquad E[(t+r-1)^{[r]}] = \frac{(k+r-1)^{[r]} \binom{N+r}{n+r}}{\binom{N}{n}}$$

where, $x^{[r]} = x(x-1) \ldots (x-r+1)$ and r is fixed. Putting $r = 1$ and 2 in (4.2), we can get after some simplification,

$$(4.3) \qquad E(t) = \frac{k(N+1)}{n+1}$$

$$(4.4) \qquad V(t) = \frac{k(N+1)(N-n)(n-k+1)}{(n+1)^2(n+2)} \quad .$$

On the other hand, considering (4.1) as probability function of $Y_{(k)}$, we get

$$(4.5) \qquad E(Y_k) = \sum_{t=k}^{N-n+k} y_{o(t)} \cdot \binom{t-1}{k-1}\binom{N-t}{n-k} \Big/ \binom{N}{n} \quad .$$

Obviously this has no simple form and so is the variance of $Y_{(k)}$.

We have mentioned earlier that (4.1) can be considered to be the probability that the k^{th} order statistic of the sample will be the t^{th} order statistic of the population. So we may want to know the most likely value of t for given k. This is given by the value of t satisfying the following relation,

$$(4.6) \qquad p_{N,n,k}(t-1) \leq p_{N,n,k}(t) \geq p_{N,n,k}(t+1)$$

i.e.

$$\frac{\binom{t-2}{k-1}\binom{N-t+1}{n-k}}{\binom{N}{n}} \leq \frac{\binom{t-1}{k-1}\binom{N-t}{n-k}}{\binom{N}{n}} \geq \frac{\binom{t}{k-1}\binom{N-t-1}{n-k}}{\binom{N}{n}}$$

This implies after some algebric manipulations,

$$(4.7) \qquad N\left(\frac{k-1}{n-1}\right) \leq t \leq N\left(\frac{k-1}{n-1}\right) + 1$$

or

$$(4.8) \qquad \frac{k-1}{n-1} \leq \frac{t}{N} \leq \frac{k-1}{n-1} + \frac{1}{N} \ .$$

As a particular case of (4.7), we can find that sample median is also the ML-estimator for the population median of finite population. This can be illustrated by the following example.

Example. Let our population size $N = 25$ and sample size $n = 9$; so that the 5^{th} largest sample value is the sample median, hence $k = 5$. So using $k = 5$ in relation (4.7), we have

$$.5N \leq t \leq .5N+1$$

or

$$12.5 \leq t \leq 13.5 \ .$$

which shows that maximum likely integer value of t is 13, but $y_{0(t)} = y_{0(13)}$ is the median of our population. Hence sample median $Y_{(5)}$ is the ML-estimator of population median $y_{0(13)}$.

□

Median unbiasedness: We have already mentioned that the sample median is the ML-estimator for the finite population median. Now, we are going to show that sample median is median unbiased when sampling is done from finite population.

Definition. If Δ is the median of a distribution and $\tilde{\Delta}$ is an estimator of Δ, then we say $\tilde{\Delta}$ is median unbiased if

$$P[\tilde{\Delta} < \Delta] = P[\tilde{\Delta} > \Delta] \quad . \tag{4.9}$$

For continuous population $P[\tilde{\Delta} = \Delta] = 0$ and hence the above probabilities are equal to 1/2. But in general, for finite population, $P[\tilde{\Delta} = \Delta] \neq 0$, so for median unbiasedness in finite population we shall consider the relation (4.9). To show median unbiasedness of sample median, we have to consider the following four cases.

(1) N and n both are odd

(2) N is odd and n is even

(3) N is even and n is odd

(4) N and n are both even.

Case 1. N and n are both odd.

Let $N = 2M+1$ and $n = 2m+1$, M and m are integers. Therefore $y_{0(M+1)}$ and $Y_{(m+1)}$ are respectively population and sample medians. Since (4.1) is probability function, we have $\sum_{t=k}^{N-n+k} P_{N,n,k}(t) = 1$. So for $k = m+1$, we have: $N-n+k = N-m$ and

$$\sum_{t=m+1}^{N-m} P[Y_{(m+1)} = y_{0(t)}] = \sum_{t=m+1}^{N-m} \binom{t-1}{m}\binom{N-t}{m}\Big/\binom{N}{n} = 1$$

or

$$\binom{m}{m}\binom{N-m-1}{m} + \binom{m+1}{m}\binom{N-m-2}{m} + \ldots + \binom{M-1}{m}\binom{M+1}{m} + \binom{M}{m}\binom{M}{m}$$

(4.10)

$$+ \binom{M+1}{m}\binom{M-1}{m} + \ldots + \binom{N-m-1}{m}\binom{m}{m} = \binom{N}{n}.$$

It is clear from (4.10), that

$$\sum_{t=m+1}^{M} P[Y_{(m+1)} = y_{0(t)}] = \sum_{t=M+2}^{N-m} P[Y_{(m+1)} = y_{0(t)}]$$

or

$$P[Y_{(m+1)} \leq y_{0(M+1)}] = P[Y_{(m+1)} \geq y_{0(M+2)}]$$

$$\vdots$$

$$P[Y_{(m+1)} < y_{0(M+1)}] = P[Y_{(m+1)} > y_{0(M+1)}] \quad .$$

Hence sample median $Y_{(m+1)}$ is median unbiased. But the situation is a little complicated in other cases, where there is no unique median.

Case 2. N is odd and n is even.

As before, let $N = 2M+1$ and $n = 2m$, M and m are integers, so that, conventionally sample median is an average of $Y_{(m)}$ and $Y_{(m+1)}$. For $k = m$, we have $N-n+k = N-m$, $n-k = m$, so that

$$(4.11)\qquad \sum_{t=m}^{N-m} P[Y_{(m)} = y_{0(t)}] = \frac{1}{\binom{N}{n}} \sum_{t=m}^{N-m} \binom{t-1}{m-1}\binom{N-t}{m} = 1 \ ,$$

and also for $k = m+1$, $N-n+k = N-m+1$, $n-k = m-1$;

$$(4.12)\qquad \sum_{t=m+1}^{N-m+1} P[Y_{(m+1)} = y_{0(t)}] = \frac{1}{\binom{N}{n}} \sum_{t=m}^{N-m+1} \binom{t-1}{m}\binom{N-t}{m-1} = 1 \ .$$

Expanding (4.11) and (4.12) as in Case 1, we find that

$$(4.13)\qquad P[Y_{(m)} < y_{0(M+1)}] + P[Y_{(m+1)} < y_{0(M+1)}]$$

$$= P[Y_{(m)} > y_{0(M+1)}] + P[Y_{(m+1)} > y_{0(M+1)}] \ ,$$

$$(4.14)\qquad P[Y_{(m)} \leq y_{0(M+1)}] = P[Y_{(m+1)} \geq y_{0(M+1)}] \ ,$$

and

$$(4.15)\qquad P[Y_{(m)} \geq y_{0(M+1)}] = P[Y_{(m+1)} \leq y_{0(M+1)}] \ .$$

If we denote $\tilde{Y} = \frac{Y_{(m)} + Y_{(m+1)}}{2}$ as the median of sample then it follows from (4.14) that, $P[\tilde{Y} < y_{0(M+1)}] = P[\tilde{Y} > y_{0(M+1)}]$ and hence median unbiasedness of $\tilde{Y}$.

Example. Let $N = 15$, $n = 6$, so that $M = 7$ and $m = 3$. The sample median $\tilde{Y} = (Y_{(3)}+Y_{(4)})/2$, the population median $\tilde{y} = y_{0(8)}$, and using (4.11),

$$\binom{N}{n} = \binom{15}{6} = 5005 \ .$$

$$P[Y_{(3)} < y_{0(8)}] = \frac{3115}{5005} \ ,$$

$$P[Y_{(3)} = y_{0(8)}] = \frac{735}{5005} ,$$

$$P[Y_{(3)} > y_{0(8)}] = \frac{1155}{5005} .$$

Again using (4.12),

$$P[Y_{(4)} < y_{0(8)}] = \frac{1155}{5005} ,$$

$$P[Y_{(4)} = y_{0(8)}] = \frac{735}{5005} ,$$

$$P[Y_{(4)} > y_{0(8)}] = \frac{3115}{5005} .$$

Therefore, $P[Y_{(3)} \leq y_{0(8)}] = P[Y_{(4)} \geq y_{0(8)}] = \frac{3850}{5005}$. And hence,

$$P[\tilde{Y} < y_{0(8)}] = P[\tilde{Y} > y_{0(8)}] .$$

Case 3. N is even and n is odd.

Let $N = 2M$ and $n = 2m+1$, M and m are integers. Let the population median be $\tilde{y}_0 = (Y_{0(M)}+y_{0(M+1)})/2$ and the sample median is $y_{(m+1)}$. For $k = m+1$, $N-n+k = N-m$. Using these values and proceeding as in Case 1, we find that

(4.16) $$P[Y_{(m+1)} \leq y_{0(M)}] = P[Y_{(m+1} \geq y_{0(M+1)}]$$

and hence

$$P[Y_{(m+1)} < \tilde{y}_0] = P[Y_{(m+1)} > \tilde{y}_0]$$

establishes that $Y_{(m+1)}$ is median unbiased.

Case 4. N and n are both even.

Let $N = 2M$ and $n = 2m$, M and n are integers. Sample and population medians are respectively, $\tilde{Y} = (Y_{(m)} + Y_{(m+1)})/2$ and $\tilde{y}_0 = (y_{0(M)} + y_{0(M+1)})/2$. Proceeding as in Case 2, we find that

$$P[Y_{(m)} \leq y_{0(M)}] + P[Y_{(m+1)} \leq y_{0(M)}] = P[Y_{(m)} \geq y_{0(M+1)}] + P[Y_{(m+1)} \geq y_{0(M+1)}] \quad , \tag{4.17}$$

$$P[Y_{(m)} \leq y_{0(M)}] = P[Y_{(m+1)} \geq y_{0(M+1)}] \quad , \tag{4.18}$$

and

$$P[Y_{(m+1)} \leq y_{0(M)}] = P[Y_{(m)} \geq y_{0(M+1)}] \quad . \tag{4.19}$$

Now,

$$P[\tilde{Y} < \tilde{y}_0] = P[\tilde{Y} < \frac{y_{0(M)} + y_{0(M+1)}}{2}] = P[\tilde{Y} \leq y_{0(M)}] = P[Y_{(m)} \leq y_{0(M)}] \quad . \tag{4.20}$$

Similarly,

$$P[\tilde{Y} > \tilde{y}_0] = P[Y_{(m+1)} \geq y_{0(M+1)}] \quad . \tag{4.21}$$

Therefore, using (4.18), (4.20) and (4.21), we get

$$P[\tilde{Y} < \tilde{y}_0] = P[\tilde{Y} > \tilde{y}_0] \quad ,$$

which establishes median unbiasedness of $\tilde{Y}$.

§4.2 CONFIDENCE INVERVALS FOR QUANTILE IN FINITE POPULATION

Let t be a fixed integer in the range $1 \leq t \leq N$. Then we can consider $y_{0(t)}$ as the $(t/N)^{th}$ quantile of the population π_N. If $G_N(y) = (\#y_{0i} \leq y,\ 1 \leq i \leq N)/N$ then we formally define the p^{th} quantile of finite population as $\sup\{y : G_N(y) < p\}$, $0 < p \leq 1$. Similarly, the sample quantile is also defined.

Confidence interval for quantile in finite population is available in Wilks (1962, p. 333). Years later Meyer (1972) and Sedransk and Meyer (1978) extensively studied and extended results on this confidence interval. For fixed $t = t'$,

$$P[Y_{(k)} \leq y_{0(t')}] = \sum_{t=k}^{t'} p_{N,n,k}(t) \quad . \tag{4.22}$$

So for fixed N,n,t' and $\gamma > 0$, there is a largest k, say k' such that

$$\sum_{t=k'}^{t'} p_{N,n,k'} \geq \gamma \quad . \tag{4.23}$$

We shall consider $Y_{(k')}$ as the best lower $100\gamma\%$ confidence limit for $y_{0(t')}$. Except for values N, n, t' and $1-\gamma$, which are uninterestingly small, such lower confidence limits can be shown to exist. Similarly, the best upper $100\gamma\%$ confidence limit for $y_{0(t')}$ is obtained by choosing the smallest k, say k'', such that

$$\sum_{t=t'}^{N-n+k''} p_{N,n,k''}(t) \geq \gamma \ . \tag{4.24}$$

For the best $100\gamma\%$ confidence interval for $y_{0(t')}$, that is the simultaneous upper and lower confidence limits, the probability

function involved here is cumbersome. Meyer (1972) has given the following expression for simultaneous confidence interval $[Y_{(k)}, Y_{(r)}]$ for $y_{0(t)}$, where

$$(4.25) \quad P[Y_{(k)} \leq y_{0(t)} \leq Y_{(r)}] = \frac{\sum_{i=o}^{t-k} \binom{t-i-1}{k-1}\binom{N-t+i}{n-k} - \sum_{i=o}^{t-r-1} \binom{t-i-2}{r-1}\binom{N-t+i+1}{n-r}}{\binom{N}{n}}.$$

A simpler form of this expression is available in Sendransk and Meyer (1978). This paper also states that in forming a confidence interval for the $(\frac{t}{N})^{th}$ quantile, the confidence coefficient for population with ties is larger than the confidence coefficient for population without ties, proof is available in Meyer (1972). In fact the confidence coefficient for the $(\frac{t}{N})^{th}$ quantile for a population without ties is the lower bound for the confidence coefficient for the comparable confidence interval for any finite population.

Confidence intervals in case of stratified sampling

Let us now consider a stratified population of q strata having strata size $N_1, \ldots, N_q$, $\sum_{i=1}^{q} N_i = N$, the population size. Let the population values in ascending order be:

$$(4.26) \quad y_{01(1)} < \cdots < y_{01(N_1)} < y_{02(1)} < \cdots < y_{02(N_2)} < \cdots < y_{0q(1)} < \cdots < y_{0q(N_q)} .$$

We have drawn a stratified random sample of size n from this population with $n = \sum n_i$, where n_i is the number of units selected at random from i^{th} stratum. Let the sample values be:

$$Y_{1(1)} < \ldots < Y_{1(n_1)} < Y_{2(1)} < \ldots < Y_{2(n_2)} < \ldots < Y_{q(1)} < \ldots < Y_{q(n_q)} \quad .$$

Let our interest be on confidence interval for $(t/N)^{th}$ quantile of the population. It is sufficient to look at the stratum which contains the i^{th} order y-value of the population. Let the t^{th} order value be in m^{th} stratum. Since a sample from each stratum is drawn independently, we shall consider for the $(t/N)^{th}$ quantile, the sample drawn from the m^{th} stratum only. Let $t' = t-(N_1 + \ldots + N_{m-1})$. Proceeding the similar way as in section one of this chapter, we can find the probability that the k^{th} order statistic of the m^{th} stratum will be equal to the t^{th} order-statistic of the population, i.e.,

$$(4.27) \qquad P[Y_{m(k)} = y_{0(t)} = y_{0m(t')}] = \binom{t'-1}{k-1}\binom{N_m - t'}{n_m - k} \Big/ \binom{N_m}{n_m} \quad ,$$

$$t' = k, k+1, \ldots, N_m - n_m + k \quad .$$

If the t^{th} order statistic of the population is not in the m^{th} stratum, but is in ℓ^{th} stratum, then

$$(4.28) \qquad P[Y_{m(k)} = y_{0(t)} = y_{0\ell(t')}] = 0 \qquad \text{for } m \neq \ell \quad \text{and}$$

$$t' = t - (N_1 + \ldots + N_{\ell-1}) \quad .$$

It appears from (4.27) and (4.28) that results for unstratified population can easily be used for stratified population with few changes in notations. Sedransk and Meyer (1978) have given results for a more general case of population. There they have not imposed the restriction (4.26) to the population and established results for population with two strata.

§4.3 JOINT DISTRIBUTION OF QUANTILES OF A SAMPLE FROM BIVARIATE FINITE POPULATION

Let our bivariate finite population be (x_{01},y_{01}), (x_{02},y_{02}), ..., (x_{0N},y_{0N}) of size N. For simplicity let us assume that there are no ties among x's and as well as among y's. Let ordered x-values be $x_{0(1)} < x_{0(2)} < \ldots < x_{0(N)}$ and ordered y-values be $y_{0(1)} < y_{0(2)} < \ldots < y_{0(N)}$. $y_{0(i)}$ does not necessarily beong to the same pair as that of $x_{0(i)}$. We have drawn a simple random sample of size n from this population. Let the sample be (X_i,Y_i), $i = 1,\ldots,n$. As above let us denote the sample ordered X-values and Y-values as $X_{(1)} < \ldots < X_{(n)}$ and $Y_{(1)} < \ldots < Y_{(n)}$ respectively. Our objective is to find the bivariate distribution of sample $(i/n)^{th}$ quantile of x and $(j/n)^{th}$ quantile of y. Let us assume $X_{(i)}$ and $Y_{(j)}$ be corresponding sample quantiles. Siddiqui (1960) has derived the joint distribution of $(X_{(i)},Y_{(j)})$ when the sample was drawn from a continuous bivariate distribution.

Distribution of $(X_{(i)},Y_{(j)})$ for a sample from finite population depends upon the nature of pairs of values (x_{0i},y_{0i}) in the population. Analogous to Siddiqui (1960), we shall introduce two new variables m_0 and M, where,

M = number of pairs (X_k,Y_k) in the sample with $X_k < X_{(i)}$ and $Y_k < Y_{(j)}$. M is a random variable which may vary from sample to sample.

m_0 = number of pairs (x_{0i},y_{0i}) in the population with $x_{0i} < x_{0(s)}$ and $y_{0i} < y_{0(t)}$. m_0 is non-random if one considers the

finite population as fixed and m_0 varies for different values of s and t.

Here, M is a dummy variable which is to be summed out from the joint distribution of $(M, X_{(i)}, Y_{(j)})$ to get our desired joint distribution of $(X_{(i)}, Y_{(j)})$.

Let $(x_0, y_{0(t)})$ and $(x_{0(s)}, y_0)$ be two units of the population which lie on the lines $y = y_{0(t)}$ and $x = x_{0(s)}$ respectively. So any one of the following five case may occur.

Case 1. $x_0 < x_{0(s)}$, $y_0 < y_{0(t)}$

Case 2. $x_0 < x_{0(s)}$, $y_0 > y_{0(t)}$

Case 3. $x_0 > x_{0(s)}$, $y_0 < y_{0(t)}$

Case 4. $x_0 > x_{0(s)}$, $y_0 > y_0(t)$

Case 5. $x_0 = x_{0(s)}$ in which case there is only one point (x_0, y_0) common to both lines $x = x_{0(s)}$ and $y = y_{0(t)}$. In such a case, the pair $(x_0, y_0) = (x_{0(s)}, y_{0(t)})$ is a measure of a unit in the population.

Let us now find the $P[M = m, X_{(i)} = x_{0(s)}, Y_{(j)} = y_{0(t)}]$ under the above mentioned different cases.

Case 1. If our population satisfies Case 1, then a possible distribution of population and sample values is given in Figure 1. For simplicity of figures, we shall assume that x-values are all positive.

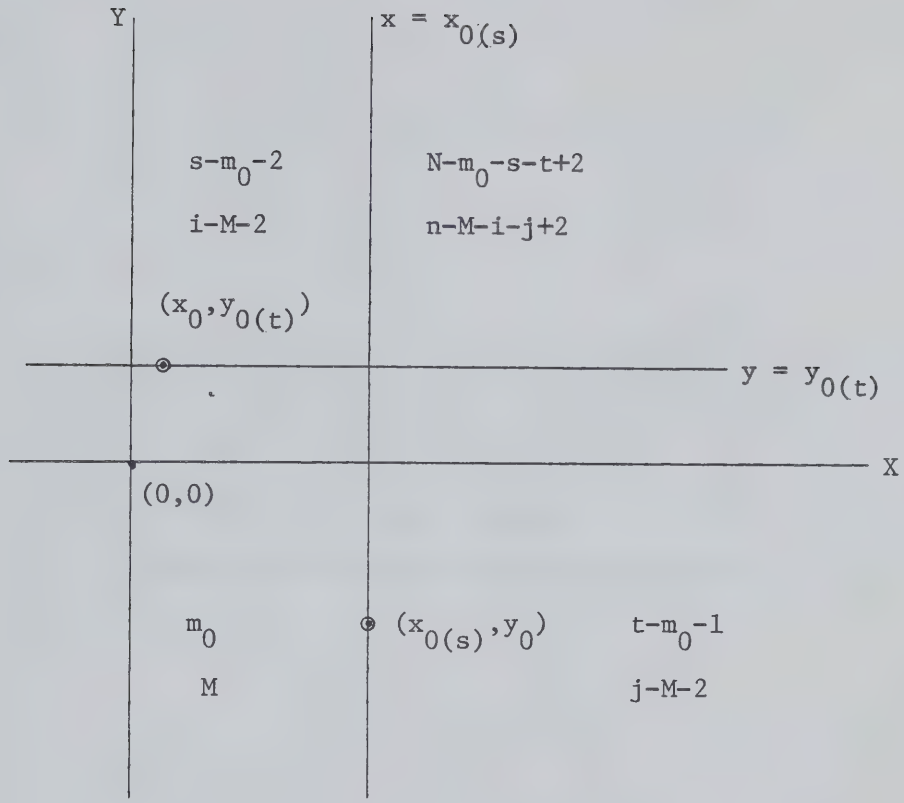

FIGURE 1

In the above figure, $N+m_0-s-t+2$ (or $n+M-i-j+2$) represents the number of pairs (x_{0i},y_{0i}) (or (X_i,Y_i)) in the population (or sample) satisfying $x_{0i} > x_{0(s)}$ and $y_{0i} > y_{0(t)}$ (or $X_i > x_{0(s)}$ and $Y_i > y_{0(t)}$). Similar meanings apply for other numbers of the figure. Points marked by ⊙ corrspond to the units of population which lead us to consider Case 1.

Since our sample is a simple random sample of size n and drawn without replacement, therefore,

$$(4.29)\qquad P_1[M = m,\ X_{(i)} = x_{0(s)}, Y_{(j)} = y_{0(t)}]$$

$$= \binom{m_0}{m}\binom{s-m_0-2}{i-m-2}\binom{t-m_0-2}{j-m-2}\binom{N+m_0-s-t+2}{n+m-i-j+2} \Big/ \binom{N}{n}$$

where,

$$m = 0,1,\ldots,m' = \min(m_0, i-2, j-2)$$

$$s = i, i+1, \ldots, N-n+i$$

$$t = j, j+1, \ldots, N-n+j \quad .$$

So (4.29) can be considered as the joint probability function of M, $X_{(i)}$ and $Y_{(j)}$ with mass points at $M = 0,1,\ldots,m'$, $X_{(i)} = x_{0(s)}$, $s = i,i+1,\ldots,N-n+i$ and $Y_{(j)} = y_{0(t)}$, $t = j,j+1,\ldots,N-n+j$. Here and for the rest of the thesis, we shall assume that for any integers p and q,

$$\binom{p}{q} = \begin{cases} 0 & \text{if } p < 0 \text{ or } p < q \text{ or } q < 0 \\ 1 & \text{if } p = q = 0 \quad . \end{cases}$$

Case 2. Proceeding as in Case 1, the configuration of the sample as well as the population values that satisfy Case 2 is given in Figure 2. Hence the required probability is

$$(4.30) \qquad P_2[M = m, X_{(i)} = x_{0(s)}, Y_{(j)} = y_{0(t)}]$$

$$= \binom{m_0}{m}\binom{s-m_0-2}{i-m-2}\binom{t-m_0-1}{j-m-1}\binom{N+m_0-s-t+1}{n+m-i-j+1} \Big/ \binom{N}{n}$$

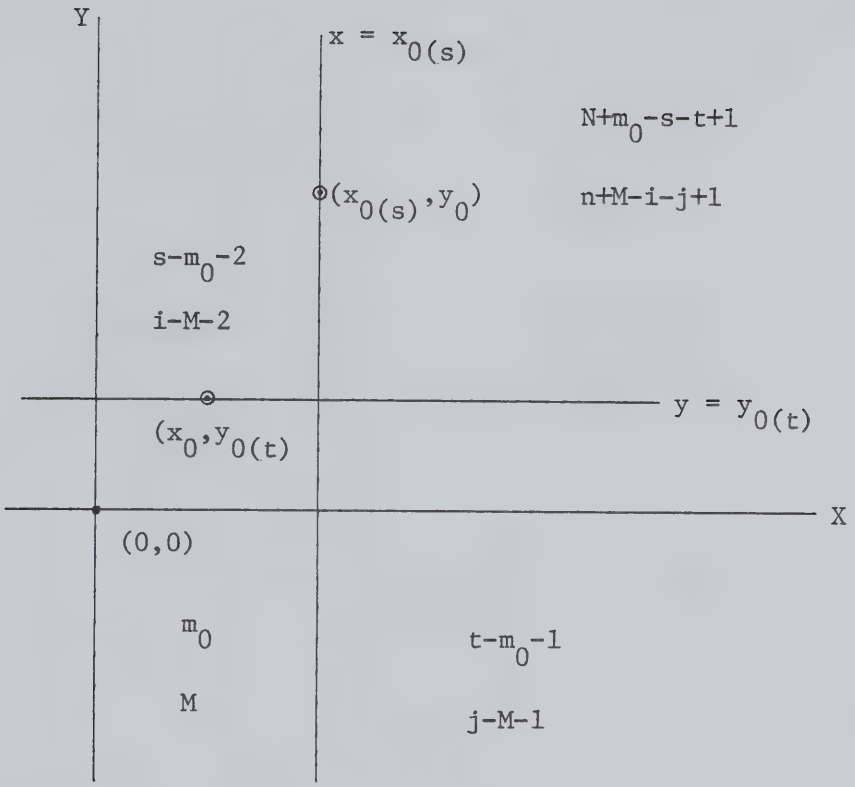

FIGURE 2

where

$m = 0,1,\ldots,m' = \min(m_0, i-2, j-1)$

$s = i, i+1, \ldots, N-n+i$

$t = j, j+1, \ldots, N-n+j$.

Case 3. The appropriate configuration for the sample as well as for the population is given in Figure 3, below.

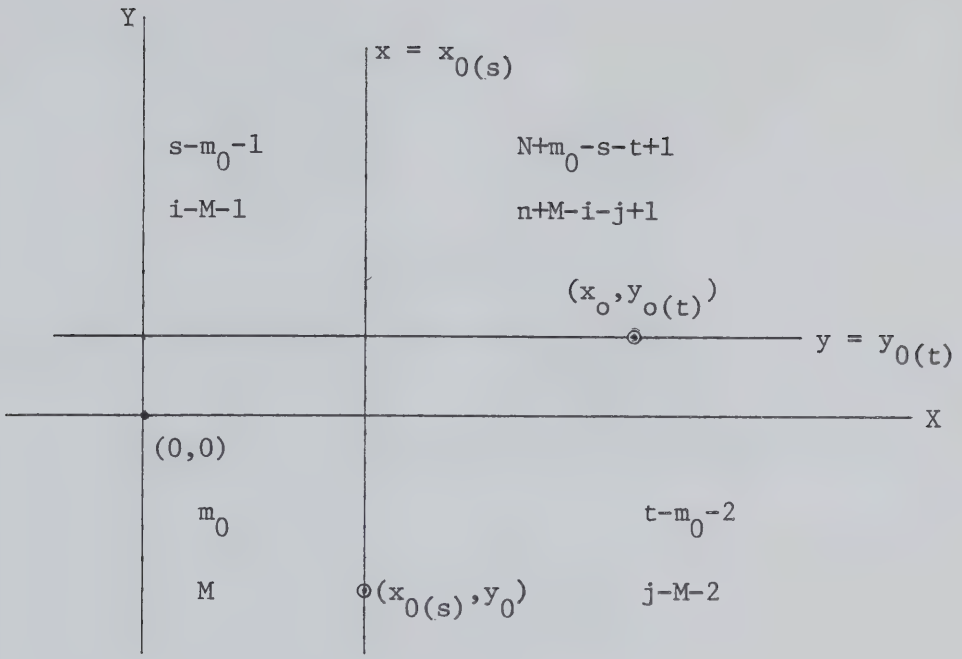

FIGURE 3

The required probability for such a configuration is

$$(4.31) \qquad P_3[M = m,\ X_{(i)} = x_{0(s)},\ Y_{(j)} = y_{0(t)}]$$

$$= \binom{m_0}{m}\binom{s-m_0-1}{i-m-1}\binom{t-m_0-2}{j-m-2}\binom{N+m_0-s-t+1}{n+m-i-j+1}\Big/\binom{N}{n}$$

where,

$$m = 0,1,\ldots,m' = \min\,(m_0,\ i-1, j-2)$$

$$s = i, i+1, \ldots, N-n+i$$

$$t = j, j+1, \ldots, N-n+j \quad .$$

Case 4. The configuration for the sample and population corresponding to Case 4, is given in Figure 4, below.

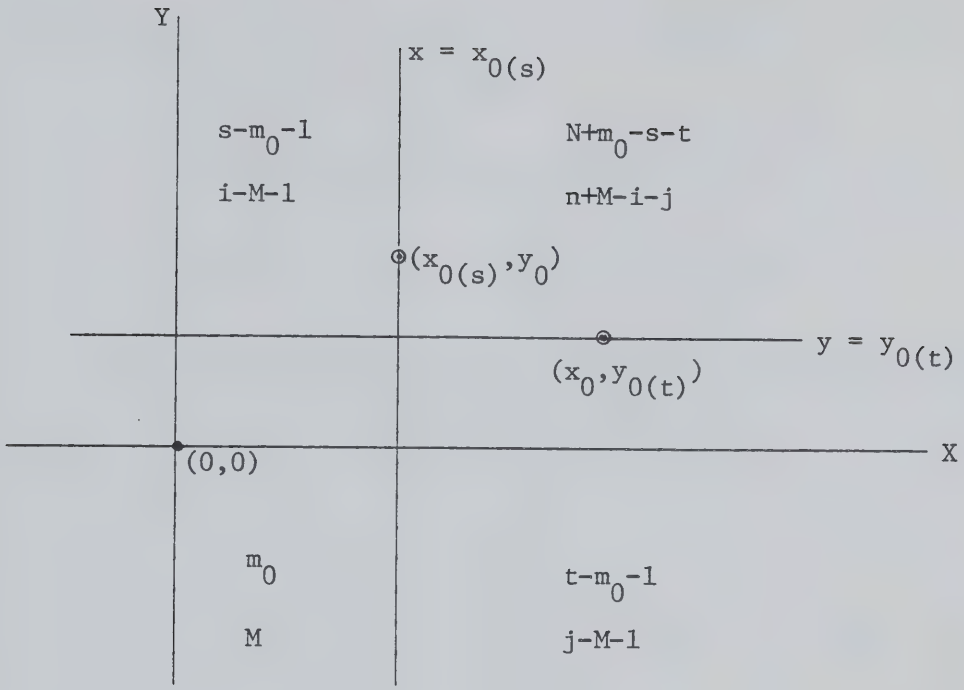

FIGURE 4

The probability for such a configuration is

$$(4.32) \qquad P_4[M = m,\ X_{(i)} - x_{0(s)},\ Y_{(j)} = y_{0(t)}]$$

$$= \binom{m_0}{m}\binom{s-m_0-1}{i-m-1}\binom{t-m_0-1}{j-m-1}\binom{N+m_0-s-t}{n+m-i-j} \Big/ \binom{N}{n}$$

where

$$m = 0,1,\ldots,m' = \min(m_0, i-1, j-1)$$

$$s = i, i+1, \ldots, N-n+i$$

$$t = j, j+1, \ldots, N-n+j \quad .$$

Case 5. A suitable configuration for Case 5 is given in Figure 5, below.

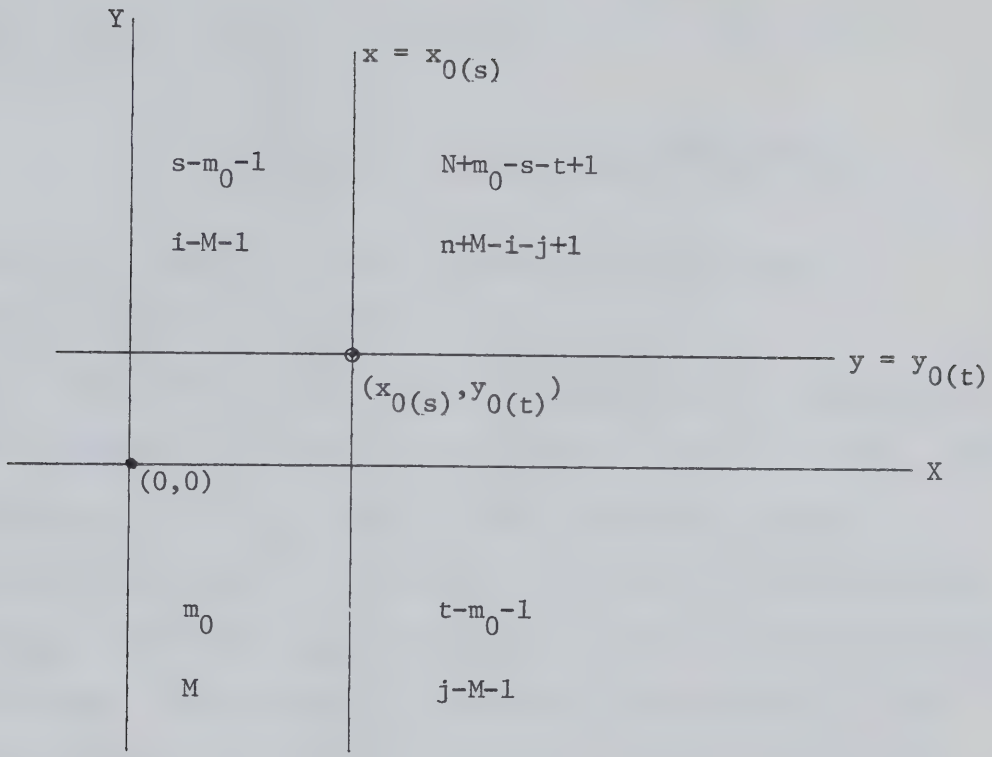

FIGURE 5

The required probability is

$$(4.33)\qquad P_5[M = m,\ X_{(i)} - x_{0(s)},\ Y_{(j)} = y_{0(t)}]$$

$$= \binom{m_0}{m}\binom{s-m_0-1}{i-m-1}\binom{t-m_0-1}{j-m-1}\binom{N+m_0-s-t+1}{n+m-i-j+1}\Big/\binom{N}{n}$$

where

$$m = 0,1,\ldots,m' = \min\,(m_0,i-1,j-1)$$

$$s = i,i+1,\ldots,N-n+i$$

$$t = j,j+1,\ldots,N-n+j\ .$$

Therefore, finally, the required probability function for $(X_{(i)},Y_{(j)})$ is

$$(4.34) \qquad P_k[X_{(i)} = x_{0(s)},\ Y_{(j)} = y_{0(t)}]$$

$$= \sum_{m=0}^{m'} P_k[M = m,\ X_{(i)} = x_{0(s)},\ Y_{(j)} = y_{0(t)}] \ ,$$

$k = 1,2,\ldots,5$ depending on which case we have at hand and

$s = i,i+1,\ldots,N-n+i$; $t = j,j+1,\ldots,N-n+j$.

Now, if we consider our finite population is a sample from a super-population with continuous distribution function $F(X,Y)$, then our m_0, $x_{0(s)}$ and $y_{0(t)}$ become random variables. In this circumstance (4.34) will be treated as a conditional distribution of $(X_{(i)},Y_{(j)})$ given m_0, $x_{0(s)}$ and $y_{0(t)}$. Since the finite population, in such a case, is a sample from a continuous distribution, so the marginal distribution of $(m_0, x_{0(s)}, y_{0(t)})$ will be as given in Siddiqui (1960).

§4.4 PREDICTION OF FINITE POPULATION QUANTILE USING AUXILIARY VARIABLES

In this section we would like to investigate the possibility of using available information on the auxiliary variable, x, to get a better estimate of population quantile of y in finite population sampling. Keeping the above objective in mind, we derived the bivariate distribution of quantiles $(X_{(i)}, Y_{(j)})$ as given by (4.34) of the preceeding section. But we could not give a simpler form to this distribution. Consequently we were unable to propose or investigate any reasonable estimator for the population quantile of y using quantiles of x.

However, if we assume certain multivarite models, namely, Model M_R or perhaps the more traditional multivariate normal distribution model at the back of our realized finite population,

then it appears that we can get a "predictor" of finite population quantiles using auxiliary information.

Here we are going to use notation developed in Chapter 2. Let $\underset{\sim}{Y} = (Y_1,\ldots,Y_N)'$ be the N-dimensional random vector giving the finite population $\underset{\sim}{y} = (y_1,\ldots,y_N)'$. Under Model M_R, let the ξ-expectation and ξ-variance of Y be

$$\mathcal{E}(Y) = X\beta \qquad \text{and} \qquad \mathcal{V}(Y) = V \ ,$$

where $\underset{\sim}{X}$ is $N\times p$ matrix, $\underset{\sim}{\beta}$ is $p\times 1$ vector and $V = \text{diag}\,(\sigma_1^2,\ldots,\sigma_N^2)$, a known $N\times N$ non-singular positive definite matrix. Let sample size $v(s) = n > p$. Let us partition $\underset{\sim}{Y}$ as:

$$\underset{\sim}{Y} = \begin{pmatrix} Y_s \\ Y_{\bar{s}} \end{pmatrix}$$

where Y_s is $n\times 1$ vector of sampled units and $Y_{\bar{s}}$ is the $(N-n)\times 1$ vector of non-sampled units. Accordingly we can partition X and V as follows:

$$X = \begin{pmatrix} X_s \\ X_{\bar{s}} \end{pmatrix}, \qquad V = \begin{pmatrix} V_s & 0 \\ 0 & V_{\bar{s}} \end{pmatrix}$$

where X_s is $n\times p$ and V_s is $n\times n$, etc.

The minimum variance unbiased estimate t_{BLU} for population total $\sum_1^N Y_i$, as given in Theorem 2.8 is

$$t_{BLU} = \ell_1' y_s + \ell_2' X_{\bar{s}} \hat{\beta}_{BLU} \tag{4.35}$$

where

$$\hat{\beta}_{BLU} = (X_s' V_s^{-1} X_s)^{-1} X_s' V_s^{-1} Y_s$$

and ℓ_1 and ℓ_2 are vectors of the form $(1,\ldots,1)'$ having dimensions n and N-n respectively.

On the other hand, if we assume our super-population model is $N(X\beta,V)$, i.e. $\underset{\sim}{Y} \sim N(X\beta,V)$, then the following theorem due to Royall (1976a) gives the maximum likelihood estimator for the population total.

Theorem 4.1. If Y has a $N(X\beta,V)$ probability distribution in which the known diagonal convariance matrix satisfies $V\ell = X\gamma$, $\ell' = (1,\ldots,1)$ for some p-vector γ, i.e. $\sigma_i^2 = \sum_{j=1}^{p} \gamma_j x_{ij}$, then when $Y_s = y_s$ is observed, the likelihood function for $t = \ell'y$ is proportional to the $N\{\hat{t}_{BLU}, \text{var}(\hat{t}_{BLU})\}$ probability density function, where $\hat{t}_{BLU}$ is given by (4.35), and

$$(4.36) \qquad \text{var}(\hat{t}_{BLU}) = \underset{\sim}{\ell}_s'\underset{\sim}{V}_s^{-1}\underset{\sim}{\ell}_s + \underset{\sim}{\ell}_s'\underset{\sim}{X}_s(\underset{\sim}{X}_s'\underset{\sim}{V}_s^{-1}\underset{\sim}{X}_s)^{-1}\underset{\sim}{X}_s'\underset{\sim}{\ell}_s \quad .$$

□

The above theorem suggests that, $\hat{t}_{BLU}$ is the best linear unbiased estimator under the normal super-population model, and that $\text{var}(t_{BLU})$ is its variance under the same model.

It is interesting to look more closely at (4.35). The term $\ell_1'y_s$ is the observed sample total and $\ell_2'\underset{\sim}{X}_s\hat{\beta}_{BLU}$ is a prediction of $\ell_2'\underset{\sim}{y}_s$, the total y_i's for non-sampled units. So, we can consider

$$(4.37) \qquad \hat{\underset{\sim}{Y}}_s = \underset{\sim}{X}_s\hat{\beta}_{BLU}$$

as the predictor of $\underset{\sim}{Y}_s$. Now if our interest is on prediction of finite population quantile, then we can combine y_s and $\hat{\underset{\sim}{y}}_s$ (the

predicted value of $\underset{\sim}{Y}_{\bar{s}}$, obtained using y_s in $\hat{\beta}_{BLU}$) to get the predicted population values of y-characters, viz., $\hat{y}' = (y_s', \hat{y}_{\bar{s}}')$. Once we have the predicted opulation $\hat{\underset{\sim}{Y}}$ at hand, we can easily sort out the required predicted quantile for finite population. A predictor of $(t/N)^{th}$ quantile of finite population will be the corresponding quantile of the above mentioned predicted population. Obviously, the predictor suggested above uses auxiliary information through $\hat{\beta}_{BLU}$.

As before, let $Y_{(t)}$ be the $(t/N)^{th}$ quantile of the finite population which we are going to predict by $\hat{Y}_{(t)}$ as per above suggestions. At this stage it is required to investigate properties such as ξ-unbiasedness and ξ-MSE of $\hat{Y}_{(t)}$. Let $\frac{t}{N} = q$, $0 < q < 1$, and Y_{sq} be the sample q^{th} quantile obtained from $\underset{\sim}{Y}_s$ without using the auxiliary information. This Y_{sq} is the commonly used predictor of corresponding population quantile $Y_{(t)}$. Now if $\mathcal{E}(\hat{Y}_{(t)} - Y_{(t)})^2 \leq \mathcal{E}(Y_{sq} - Y_{(t)})^2$ for any sample $s \in \mathcal{A}_n$, then we shall have our proposed predictor $\hat{Y}_{(t)}$ at least as good as the predictor Y_{sq} and if the strict inequality holds for some s, then our predictor $\hat{Y}_{(t)}$ will be better than the predictor Y_{sq} under the Model G_{MR} or multivariate normal super-population model.

To study these properties we need moments or distribution of $\hat{Y}_{(t)}$ which at present we are unable to find. If we assume the above mentioned multivariate normal super-population, then the marginal distribution of $\hat{\underset{\sim}{Y}}_{\bar{s}}$ is also multivariate normal. But the joint distribution of $\hat{Y}' = (Y_s', \hat{Y}_{\bar{s}}')$ is multivariate normal with singular variance-covariance matrix with rank n. Obviously $\hat{Y}_i$'s are no longer independent as well as identically distributed rather

the distribution depends on the columns of $\underset{\sim}{X}$. Exact distribution or moments of order statistics for dependent variates have been studied by Young (1967), Greig (1967) and Afonja (1972). But there they considered the parent distribution is either exchangeable or has equal correlation among the variates or has non-singular variance-covariance matrix. In survey sampling both N and n are usually very large. So, asymptotic properties (as $N,n \to \infty$) may be of some interest. Some general results on the asymptotic behaviour of function of order statistics with different mixing types of dependence are available in Gastwirth and Rubin (1975) and Mehra and Rao (1975). But these types of dependence apparently do not correspond to the nature of dependence we have in our $\hat{\underset{\sim}{Y}}$. So, for developing useful properties of our estimator $\hat{Y}_{(t)}$, further study will be required on the exact and asymptotic distribution of order statistics and their functions where the sample is drawn from the population with singular variance-covariance matrix.

CHAPTER V

ASYMPTOTIC RESULTS FOR SAMPLES FROM FINITE POPULATION

§5.1 INTRODUCTION

It is a common practice in survey sampling to use the central limit theorem for large populations and through this central limit theorem we use standard tests for testing hypothesis concerning finite population parameters. Rosen (1964) has given a systematic analytic basis for asymptotic behavior of our statistics based on sampling from finite population. There are also some earlier works in this area, namely, Erdos and Renyi (1959) and Hajek (1960). Recent works on the asymptotic behavior of order statistics and quantiles of a sample from finite population are Jha (1975) and Singh (1980). In this chapter we shall state some of these results and then we shall derive the asymptotic bivariate distribution of sample quantiles.

§5.2 SOME ASYMPTOTIC RESULTS

Usually we consider our population, $\pi = (y_{01},\ldots,y_{0N})$, as a finite set of fixed numbers and the sample of size n drawn from this population is $y_1,\ldots,y_n$. When the sampling is done without replacement, sample observations become correlated due to sampling. Although this dependency can be ignored for sufficiently large population size N, for samples drawn without replacement the conventional limit procedure for independent observations as $n \to \infty$ does not have any meaning. The population will be exhausted after a finite number of drawings. So many authors

considered a double sequence of random variables as follows:

$$\begin{array}{ll} Y_{11}, Y_{12}, \ldots, Y_{1n_1} & \text{is a random sample from } \pi_1 \\ \cdot \quad \cdots \quad \cdot & \\ \cdot \quad \cdots \quad \cdot & \\ \cdot \quad \cdots \quad \cdot & \\ Y_{k1}, Y_{k2}, \ldots, Y_{k,n_k} & \text{is a random sample from } \pi_k \ . \end{array}$$

They considered the limiting behaviour of statistics based on the sequence $\{\pi_k\}_1^\infty$ of population and assumed that the k^{th} population size $N_k \to \infty$ as $k \to \infty$. Let

$$(5.1) \qquad \mu_k = \frac{1}{N_k} \sum_{j=1}^{N_k} y_{0kj} \qquad \text{and} \qquad \sigma_k^2 = \frac{1}{N_k - 1} \sum_{j=1}^{N_k} (y_{0kj} - \mu_k)^2$$

respectively k^{th} population mean and variance. Let F_k be the k^{th} population distribution function obtained by giving weight $1/N_k$ to each element of π_k. F_k is assumed to be right continuous. The centered distribution function $F_k^c(y)$ is defined by $F_k^c(y) = F_k(y-\mu_k)$. Let $\bar{Y}_n^{(k)} = (Y_{k1} + \ldots + Y_{kn})/n$. Then the following two theorems establish the convergence of sample mean in finite population sampling.

Theorem 5.1. (Rosen, 1964): Let $\{\pi_k\}_1^\infty$ be a sequence of populations. A sufficient condition for $\bar{Y}_n^{(k)} - \mu_k$ to converge almost surely to 0 for the sample size sequence (SSS) $\{n_k\}_1^\infty$ is that,

$$(5.2) \qquad \lim_{k \to \infty} \sigma_k^2 \left(\frac{1}{n_k} - \frac{1}{N_K} \right) = 0 \ .$$

If $\{\pi_k\}_1^\infty$ satisfies

$$\lim_{A \to \infty} \sup \int_{|y|>A} y^2 dF_k^c(y) = 0 \tag{5.3}$$

then condition (5.2) is necessary for $\bar{Y}_n^{(k)} - \mu_k$ to converge in probability to 0 for SSS $\{n_k\}_1^\infty$.

□

Theorem 5.2. (Rosen, 1964). Necessary and sufficient conditions that $\bar{Y}_n^{(k)} - \mu_k$ converges to 0 with probability 1 for every SSS $\{n_k\}_1^\infty$ with $n_k \to \infty$ when $k \to \infty$ is that $\{\pi_k\}_1^\infty$ satisfies,

$$\lim_{A \to \infty} \sup_k \int_{|y|>A} |y| dF_k(y) = 0 \ . \tag{5.4}$$

□

Central limit theorem for finite population has been studied by Erdös and Renyi (1959) and Hajek (1960, 1961). The following theorem, due to Hajek (1960), gives a necessary and sufficient condition for sample total to be asymptotically normally distributed. As mentioned in Chapter 1, let $\mathcal{U}_k = \{1,2,\ldots,N_k\}$ be the label set of the k^{th} population (π_k) units. Let s_k be a simple random sample of size n_k from $\mathcal{U}_k$. So that the k^{th} sample total $n_k \bar{Y}^{(k)} = \sum_{i \in s_k} Y_{ki}$ has mean and variance equal to $n_k \mu_k$ and $((N_k - n_k)/N_k) n_k \sigma_k^2 = D_k$, respectively, where μ_k and σ_k^2 are as defined in (5.1).

Theorem 5.3. (Hajek, 1960). Let $s_{k\tau}$ be the subset of elements of $\mathcal{U}_k$ on which the inequality

$$|y_{0ki} - \mu_k| > \tau \sqrt{D_k} \tag{5.5}$$

holds, where D_k is the variance of k^{th} sample total, $n_k\bar{Y}^{(k)}$. Suppose $n_k \to \infty$ and $(N_k - n_k) \to \infty$.

Then the random variable $n_k\bar{Y}^{(k)}$ has asymptotically normal distribution with parameter $(n_k\mu_k, D_k)$ if and only if

$$\lim_{k \to \infty} \frac{\sum_{i \in s_{k\tau}} (y_{0ki} - \mu_k)^2}{\sum_{i=1}^{N_k} (y_{0ki} - \mu_k)^2} = 0 , \quad \text{for any } \tau > 0. \tag{5.6}$$

□

Estimation of quantiles is usually considered with hardly any restriction concerning the distribution. In the first situation an efficient estimator for the unknown quantile can be derived from the efficient estimator of the unknown parameter. In the second case, the natural estimator, namely the sample quantile, cannot be beaten, Reiss (1980). We shall now discuss asymptotic behaviour of sample quantiles. For this, we need the concept of empirical distribution function $G(t,n)$ corresponding to the sample $Y_1,\ldots,Y_n$ from π, which is defined as

$$G(t,n) = \frac{1}{n} \sum_{j=1}^{n} I(t - Y_j) , \qquad -\infty < t < \infty , \tag{5.7}$$

where $I(\cdot)$ is indicator function with

$$\begin{aligned} I(u) &= 0 \quad \text{if} \quad u < 0 \\ &= 1 \quad \text{if} \quad u \geq 0 \quad . \end{aligned} \tag{5.8}$$

If no complexity arises then we shall use $G(t)$ for $G(t,\mathbf{n})$.

Let $0 < p < 1$. Then the p^{th} quantile of a distribution function $F(t)$ is defined as supremum over the t-values for which $F(t) < p$. Analogously, we define the empirical p^{th} quantile corresponding to a sample of size n from π as the surpremum over the t-values for which $G(t,n) < p$. The following theorem, due to Rosen (1964), gives the asymptotic behaviour of empirical p^{th} quantile.

Theorem 5.4. Let $Y(p,n_k)$ be the empirical p^{th} quantile in a sample of size n_k from π_k, $k = 1,2,\ldots$. We assume that there is a continuous distribution function $F(t)$ such that

$$\lim_{k \to \infty} \sqrt{n_k} \sup_t \left| F_{\pi_k}(t) - F(t) \right| = 0 \tag{5.9}$$

and, furthermore, that $F(t)$ is continuous and positive in a vicinity of the p^{th} quantile η_p of $F(t)$. Now, if $\lim_{k \to \infty} n_k = \infty$ and

$$\overline{\lim} \frac{n_k}{N_k} < 1$$

then for every real α,

$$\lim_{k \to \infty} P\left[\frac{F'(\eta_p)Y(p,n_k) - \eta_p}{\sqrt{p(1-p)\left(\frac{1}{n_k} - \frac{1}{N_k}\right)}} \leq \alpha \right] = \frac{1}{2\pi} \int_{-\infty}^{\alpha} e^{-x^2/2}\, dx \quad . \tag{5.10}$$

□

This work of Rosen was later extended by Singh (1980). Singh (1980) has shown that after proper normalization, the weak limit of the process $G_k^{-1}(t)$ is W^o, where W^o is a Brownian bridge on $D[0,1]$, the space of all right continuous functions on $[0,1]$ having left hand limit (for details on the space $D[0,1]$ please see Billingsley (1968)).

§5.3 ASYMPTOTIC BIVARIATE DISTRIBUTION OF SAMPLE QUANTILES

In Chapter 4, we have derived the joint distribution of $(X_{(i)}, Y_{(j)})$ for a sample from a bivariate distribution of finite population. There we considered $X_{(i)}$ and $Y_{(j)}$ as sample $(i/n)^{th}$ quantile of x-values and $(j/n)^{th}$ quantile of y-values respectively and depending on the population configuration, we derived five different forms of probability functions for $(X_{(i)}, Y_{(j)})$. In this section we shall study asymptotic behaviour of those distributions.

We shall first consider the probability function under Case 1 of Chapter 4, (relation (4.29) with s replaced by r). Let us assume that N is so large such that m_0, $r-m_0-2$, $t-m_0-2$ and $N+m_0-r-t+2$ are also sufficiently large for applying the following approximations:

$$(5.11)\qquad \binom{m_0}{m} = \frac{m_0!}{m!(m_0-m)!} = \frac{m_0(m_0-1)\ \ldots\ (m_0-m+1)}{m!}$$

$$= \frac{m_0^m \{1 \cdot (1-\frac{1}{m_0})(1-\frac{2}{m_0})\ \ldots\ (1-\frac{m-1}{m_0})\}}{m!}$$

$$\simeq \frac{m_0^m}{m!}, \quad \text{for large } m_0 .$$

Similarly,

$$(5.12)\qquad \binom{r-m_0-2}{i-m-2} \simeq \frac{(r-m_0-2)^{i-m-2}}{(i-m-2)!},$$

$$(5.13)\qquad \binom{t-m_0-2}{j-m-2} \simeq \frac{(t-m_0-2)^{j-m-2}}{(j-m-2)!},$$

$$(5.14) \qquad \binom{N+m_0-r-t+2}{n+m-i-j+2} \approx \frac{(N+m_0-r-t+2)^{n+m-i-j+2}}{(n+m-i-j+2)!} \quad ,$$

$$(5.15) \qquad \binom{N}{n} \approx \frac{N^n}{n!} \quad .$$

Case 1 implies that units $(x_{0(r)},y_0)$ and $(x_0,y_{0(t)})$ are both in the sample s, and $x_0 < x_{0(r)}$, $y_0 < y_{0(t)}$. Probability that these two particular units will be in the sample is,

$$(5.16) \qquad \frac{1}{N(N-1)} = \Pr[\{(x_{0(r)},y_0), (x_0,y_{0(t)})\} \in s] \quad .$$

For large N, (5.16) can be approximated by

$$(5.17) \qquad \frac{1}{N^2} = \Pr[(x_{0(r)},y_0) \in s] \cdot P[(x_0,y_{0(t)}) \in s] \quad .$$

Let us denote $\lim_{N \to \infty} r/N = \alpha$, $\lim_{N \to \infty} t/N = \beta$ and the joint distribution function of (X,Y) as $N \to \infty$ by $F(X,Y)$, which we shall assume continuous. So for large N, we can express the right-hand side of (5.17) as

$$\Pr[x_\alpha < X < x_\alpha + \Delta x,\ y_0 < y_\beta] \cdot \Pr[x_0 < x_\alpha,\ y_\beta < Y < y_\beta + \Delta y] \quad .$$

But

$$(5.18) \qquad \Pr[x_\alpha < X < x_\alpha + \Delta x,\ y_0 < y_\beta] =$$

$$= \frac{\Pr[X < x_\alpha + \Delta x,\ y_0 < y_\beta] - \Pr[X < x_\alpha,\ y_0 < y_\beta]}{\Delta x} \cdot \Delta x$$

$$= \left.\frac{\partial F(x,y)}{\partial x}\right|_{(x_\alpha,y_\beta)} dx \qquad \text{as} \qquad \Delta x \to 0 \quad .$$

Similarly,

$$(5.19)\qquad \Pr[x_0 < x_\alpha,\ y_\beta < Y < y_\beta + \Delta y] = \left.\frac{\partial F(x,y)}{\partial y}\right|_{(x_\alpha, y_\beta)} dy \quad \text{as } \Delta y \to 0 \ .$$

So, finally considering $P_1[M = m,\ X_{(i)} = x_{0(r)},\ Y_{(j)} = y_{0(t)}]$ of finite case equal to $P_1[m, x_\alpha, y_\beta]\, dxdy$ as $N \to \infty$, we get by using (5.11)-(5.19) in (4.29)

$$(5.20)\qquad P_1[m, x_\alpha, y_\beta]dxdy = \frac{n!}{m!(i-m-2)!(j-m-2)!(n+m-i-j+2)!}$$

$$\times \left(\frac{m_0}{N}\right)^m \left(\frac{r-m_0-2}{N}\right)^{i-m-2} \left(\frac{t-m_0-2}{N}\right)^{j-m-2}$$

$$\times \left(\frac{N+m_0-r-t+2}{N}\right)^{n+m-i-j+2} \cdot \frac{\partial F}{\partial x}\,\frac{\partial F}{\partial y}\, dxdy \ ,$$

where partial derivatives are evaluated at (x_α, y_β). Let as $N \to \infty$,

$$\frac{m_0}{N} = p_1$$

$$\frac{r-m_0-2}{N} = p_2$$

$$\frac{t-m_0-2}{N} = p_3$$

$$\frac{N+m_0-r-t+2}{N} = p_4 \ .$$

So that, we can write (5.20) as:

$$(5.21)\qquad P_1[m, x_\alpha, y_\beta]dxdy = \frac{n!\, p_1^m\, p_2^{i-m-2}\, p_3^{j-m-2}\, p_4^{n+m-i-j+2}}{m!(i-m-2)!(j-m-2)!(n+m-i-j+2)!} \times \frac{\partial F}{\partial y}\,\frac{\partial F}{\partial y}\,dxdy \ .$$

Our (5.21) is exactly the same as in (3.1) of Siddiqui (1960). Similarly we can approximate for Case 2, Case 3, Case 4 and Case 5. Hence our bivariate distribution of $(X_{(i)}, Y_{(j)})$ for finite population, as $N \to \infty$, conforms to the bivariate distribution of $(X_{(i)}, Y_{(j)})$ for continuous population as derived by Siddiqui (1960).

REFERENCES

AFONJA, B. (1972). The moments of the maximum of correlated normal and t-variates. J.R. Statist. Soc. B. 34, 251-262.

BARNARD, G.A. (1971). Discussion of paper by V.P. Godambe and M.E. Thompson. J.R. Statist. Soc. B. 33, 276-378.

BASU, D. (1969). Role of the sufficiency and likelihood principles in sample survey theory. Sankhya, A. 31, 441-454.

BASU, D. (1971). An essay on the logical foundations of survey sampling, part one. In V.P. Godambe and D.A. Sprott, Eds., Foundations of Statistical Inference. Toronto: Holt, Rinehart and Winston, 203-242.

BASU, D. (1978). Relevance of randomization in data analysis (with discussion), in Survey Sampling and Measurement, ed. N.K. Namboodiri, New York, Academic Press, 267-339.

BASU, D. (1980). Randomization analysis of experimental data: The Fisher randomization test (with discussion). J. Amer. Statist. Asso. 75, 575-595.

BILLINGSLEY, P. (1968). Convergence of Probability Measures. New York: Wiley.

BREWER, K.R.W. (1963). Ratio estimation and finite population: Some results deducible from the assumption of an underlying stochastic process. Aust. J. Statist. 5, 93-105.

CASSEL, C.M., SARNDAL, C.E., and WRETMAN, J.H. (1976). Some results on generalized regression estimation for finite populations. Biometrika, 63, 615-620.

CASSEL, C.M., SARDNAL, C.E., and WRETMAN, J.H. (1977). Foundations of Inference in Survey Sampling. Wiley-Interscience.

COCHRAN, W.G. (1939). The use of analysis of variance in enumeration by sampling. J. Amer. Statist. Ass. 34, 492-510.

COCHRAN, W.G. (1946). Relative accuracy of systematic and stratified random samples for a certain class of populations. Ann. Math. Statist. 17, 164-177.

COCHRAN, W.G. (1977). Sampling Techniques, 3rd ed., New York: Wiley.

DEMING, W.E. and STEPHAN, F. (1941). On the interpretation of census as samples. J. Amer. Statist. Ass. 36, 45-49.

ERDÖS, P. and RENYI, A. (1959). On central limit theorem for samples from a finite population. Publ. Math. Inst. Hung. Acad. Sci., 4, 49-61.

ERICSON, W.A. (1965). Optimum stratified sampling using prior information. J. Amer. Statist. Ass. 60, 750-771.

ERICSON, W.A. (1969a). Subjective Bayesian models in sampling finite populations. J. Roy. Statist. Soc. B. 31, 195-224.

ERICSON, W.A. (1969b). Subjective Bayesian models in sampling finite populations: Stratification. In N.L. Johnson and H. Smith, Eds., New Developments in Survey Sampling. New York: Wiley-Interscience, 326-357.

FORMAN, E.K. and BREWER, K.R.W. (1971). The efficient use of supplementary information in standard sampling procedures. J. Roy. Statist. Soc. B. 33, 391-400.

FULLER, W.A. (1970). Simple estimators for the mean of skewed populations. Tech. Report. Iowa State University.

GASTWIRTH, J.L. and RUBIN, H. (1975). The behavior of robust estimators on dependent data. Ann. Statist. 3, 1070-1100.

GODAMBE, V.P. (1966). A new approach to sampling from finite populations I. J. Roy. Statist. Soc. B. 28, 310-328.

GODAMBE, V.P. (1968). Bayesian sufficiency in survey-sampling. Ann. Math. Statist. 20, 363-373.

GODAMBE, V.P. and JOSHI, V.M. (1965). Admissibility and Bayers estimation in sampling from finite populations, I. Ann. Math. Statist., 36, 1707-1722.

GODAMBE, V.P. and THOMPSON, M.E. (1971). Bayes, fiducial and frequency aspects of statistical inference in regression analysis in survey sampling. J. Roy. Statist. Soc. B. 38, 361-390.

GODAMBE, V.P. and THOMPSON, M.E. (1973). Estimation in sampling theory with exchangeable prior distributions. Ann. Statist. 1, 1212-1221.

GREIG, M. (1967). Extremes in a random assembly. Biometrika. 54, 273-282.

HACKING, I. (1965). Logic of Statistical Inference. London, Cambridge University Press.

HA'JEK, J. (1960). Limiting distributions in simple random sampling from a finite population. Publ. Math. Inst. Hung. Acad. Sci., 5, 361-374.

HA'JEK, J. (1961). Some extension of the Wald-Wolfowitz-Noether theorem. Ann. Math. Statist., 32, 506-523.

HANSEN, M.H. and HURWITZ, W.N. (1943). On the theory of sampling from finite populations. Ann. Math. Statist., 14, 333-362.

HARTLEY, H.O. and RAO, J.N.K. (1968). A new estimation theory for sample surveys. Biometrika, 55, 547-557.

HARTLEY, H.O. and RAO, J.N.K. (1969). A new estimation theory for sample surveys, II. In N.L. Johnson and H. Smith, Eds., New Developments in Survey Sampling. New York: Wiley-Interscience, 147-164.

HARTLEY, H.O. and SIELKEN, R.L. Jr., (1975). A super-population viewpoint for finite population sampling. Biometrics, 31, 411-422.

HOLT, D. (1975). A generalization of balanced sampling. Sankhya, 37, C., 199-203.

HORVITZ, D.G. and THOMPSON, D.J. (1952). A generalization of sampling without replacement from finite universe. J. Amer. Statist. Ass. 47, 663-685.

JESSEN, R.J. (1942). Statistical Investigation of a sample survey for obtaining farm facts. Iowa Agr. Exp. Sta., Res. Bull. No. 304.

JHA, V.D. (1975). Asymptotic distribution concerning discrete order statistics. The Mathematics Eduction, Vol. 9, No. 2, 29-32.

KALBFLEISCH, J.D. and SPROTT, D.A. (1969). Applications of likelihood and fiducial probability to sampling finite populations. In N.L. Johnson and H. Smith, Eds., New Developments in Survey Sampling. New York: Wiley-Interscience, 358-389.

KEMPTHORNE, O. (1969). Some remarks on statistical inference in finite sampling. In N.J. Johnson and H. Smith, Eds. New Developments in Survey Sampling. New York: Wiley-Interscience, 671-695.

KOLEHMAINEN, O. (1981). Bayesian models in estimating the total of a finite population: Towards a general theory. Scand. J. Statist., 8, 27-32.

MADOW, W.G. and MADOW, L.H. (1944). On the theory of systematic sampling. Ann. Math. Statist., 15, 1-24.

MAHALANOBIS, P.C. (1944). On large scale sample surveys. Roy. Soc. Phil. Trans. B. 231, 329-451.

MEHRA, K.L. and RAO, M.S. (1975). On functions of order statistics for mixing processes. Ann. Statist., 3, 874-883.

MEYER, J.S. (1972). Confidence intervals for quantiles in stratified random sampling. Unpublished Ph. D. dissertation, Iowa State University Library, Ames, Iowa.

MUKHOPADHYAY, P. (1977). Robust estimators of finite population total under certain linear regression models. Sankhya, 39 C, 71-87.

NEYMAN, J. (1971). Discussion on paper by R.M. Royall. In V.P. Godambe and D.A. Sprott, Eds., Foundations of Statistical inference. Toronto: Holt, Rinehart and Winston, 276-278.

RAJ, D. (1958). On the relative accuracy of some sampling techniques. J. Amer. Statist. Ass., 53, 98-101.

RAMAKRISHNAN, M.R. (1970). Optimum estimators and strategies in survey sampling. Ph.D. Thesis, Indian Statistical Institute.

RAO, C.R. (1971). Some aspects of statistical inference in problems of sampling from finite populations. In V.P. Godambe and D.A. Sprott, Eds., Foundations of Statistical Inference. Toronto: Holt, Rinehart and Winston, 197-202.

RAO, J.N.K. (1975). On the foundations of survey sampling. In J.N. Shrivastara, Eds., A Survey of Statistical Design and Linear Models. The Hague: North Holland, 489-505.

RAO, J.N.K. and BELLHOUSE, D.R. (1978). Optimal estimation of a finite population mean under generalized random permutation models. Carleton Mathematical Series, No. 151.

REISS, R.D. (1980). Estimation of quantiles in certain non-parametric models. Ann. Statist. 8, 87-105.

RINGER, L.R., JENKINS, O.C. and HARTLEY, H.O. (1972). Roof estimators for the mean of skewed distributions. J. Amer. Statist. Ass., 68, 414-419.

ROSE'N, B. (1964). Limit theorems for sampling from finite populations. Arkiv for Mathematik., 5, 383-424.

ROYALL, R.M. (1968). An old approach to finite population sampling theory. J. Amer. Statist. Ass., 63, 1269-1279.

ROYALL, R.M. (1970a). Finite population sampling on labels in estimation. Ann. Math. Statist., 41, 1774-1779.

ROYALL, R.M. (1970b). On finite population sampling theory under certain linear regression models. Biometrika, 57, 377-387.

ROYALL, R.M. (1971). Linear regression models in finite population sampling theory. In V.P. Godambe and D.A. Sprott, Eds., Foundation of Statistical Inference. Toronto: Holt, Rinehart and Winston, 259-274.

ROYALL, R.M. (1976a). Likelihood functions in finite population sampling theory. Biometrika, 63, 605-614.

ROYALL, R.M. (1976b). The linear least squares prediction approach to two-stage sampling. J. Amer. Statist. Ass., 71, 657-664.

ROYALL, R.M. and DUMBERLAND, W.G. (1978a). Variance estimation in finite population sampling. J. Amer. Statist. Asso., 73, 351-358.

ROYALL, R.M. and CUMERLAND, W.G. (1978b). An empirical study of prediction theory in finite population sampling: Simple random sampling and ratio estimator, (with discussion). In Survey Sampling and Measurement, ed. N.K. Namboodiri, New York: Academic Press, 293-335.

ROYALL, R.M. and CUMBERLAND, W.G. (1981). An empirical study of the ratio estimator and estimators of the variance, (with discussions). J. Amer. Statist. Asso., 76, 66-88.

ROYALL, R.M. and HERSON, J. (1973a). Robust estimation in finite populations I. J. Amer. Statist. Ass., 68, 880-889.

ROYALL, R.M. and HERSON, J. (1973b). Robust estimation in finite populations II: Stratification on a size variable. J. Amer. Statist. Ass., 68, 890-893.

SARNDAL, C.E. (1978). Design-based and model-based inference in survey sampling. Scand. J. Statist., 5, 27-52.

SCOTT, A.J., BREWER, K.R.W. and HO, E.W.H. (1978). Finite population sampling and robust estimation. J. Amer. Statist. Ass., 73, 359-361.

SCOTT, A.J. and SMITH, T.M.F. (1969). Estimation in multistage surveys. J. Amer. Statist. Ass., 64, 830-840.

SCOTT, A.J. and SMITH, T.M.F. (1974). Linear super-population models in survey sampling. Sankhya, 36. C., 143-146.

SEDRANSK, J. and MEYER, J. (1978). Confidence intervals for the quantiles of a finite population: Simple random and stratified simple random sampling. J. Roy. Statist. Soc., B., 40, 239-252.

SIDDIQUI, M.M. (1960). Distribution of quantiles in samples from a bivariate population. J. Res. Nat. Bu. Standards, B., 64B, 145-150.

SINHA, B.K. (1976). On balanced sampling schemes. Calcutta Statist. Ass. Bull., 25, 129-138.

SINGH, K. (1980). A note on sample quantile process of finite populations. Austral. J. Statist., 22, 358-363.

SINGH, P. and GARG, J.N. (1979). On balanced random sampling. Sankhya, 41. C., 60-68.

SMITH, H.F. (1938). An empirical law governing soil heterogeneity. J. Agr. Sci., 28, 1-23.

SOLOMAN, H. and ZACKS, S. (1970). Optimal design of sampling from finite populations: A critical review and indication of new research areas. J. Amer. Statist. Ass., 65, 653-677.

TALLIS, G.M. (1978). Note on robust estimation in finite populations. Sankhya, 40, C., 136-138.

THOMSEN, I. (1978). Discussion of Carl-Eric Sarndal's paper. Scand. J. Statist., 5, 43-45.

WILKS, S.S. (1962). Mathematical Statistics. John Wiley and Sons, Inc., New York.

YOUNG, D.H. (1967). Recurrence relations between the P.D.F.'s of order statistics of dependent variables, and some applications. Biometrika, 54, 283-292.

ZACKS, S. (1969). Bayes sequential designs for sampling finite populations. J. Amer. Statist. Ass., 64, 1342-1349.